



Modelação dos padrões da criminalidade no Rio Grande do Sul,
Brasil

Fernanda Krás Borges Barth

Dissertação apresentada como requisito parcial para
obtenção do grau de Mestre em Ciência e Sistemas
de Informação Geográfica

NOVA Information Management School

Modelação dos padrões da criminalidade no Rio Grande do Sul, Brasil

Dissertação orientada por
Professora Doutora Ana Cristina Costa

Outubro de 2019

DECLARAÇÃO DE ORIGINALIDADE

Declaro que o trabalho contido neste documento é da minha autoria e não de outra pessoa. Toda a assistência recebida de outras pessoas está devidamente assinalada e é efetuada referência a todas as fontes utilizadas (publicadas ou não).

O trabalho não foi anteriormente submetido ou avaliado na NOVA Information Management School ou em qualquer outra instituição.

Lisboa, 22/10/2019

Fernanda Krás Borges Barth

[a versão assinada pelo autor encontra-se arquivada nos serviços da NOVA IMS]

AGRADECIMENTOS

Meus sinceros agradecimentos à professora orientadora Ana Cristina Costa, que foi sempre atenciosa, interessada e disponível. Muito obrigada pela paciência, correções e conselhos, sempre respondendo rapidamente. Obrigada por me ensinar e contagiar com o entusiasmo em geoestatística.

À minha mãe e minha avó, por sempre me apoiarem incondicionalmente, sempre incentivando meus estudos.

Ao Max Denig, por estar sempre presente, pela ajuda em estatística, além de todo o apoio.

Agradeço também a todos os meus amigos, em especial Vinicius Jacques, Matheus Kummel, Renata Quevedo e Gabriela Viana, cujo apoio e amizade foram muito importantes.

Muito obrigada a todos!

RESUMO

O Brasil possui altas taxas de criminalidade, sendo a taxa de homicídios uma das mais altas do mundo. Apesar de permanecer com valores baixos em relação aos restantes estados, o Rio Grande do Sul teve um crescimento preocupante nas taxas de criminalidade nos últimos anos. Este trabalho tem por objetivo investigar os padrões espaciais da criminalidade nos municípios do Rio Grande do Sul, sobretudo da taxa de roubo e de homicídio, assim como os processos que influenciam esses padrões. Para tal, foram investigados diferentes modelos de regressão, tanto lineares (OLS – *Ordinary Least Squares* e GWR – *Geographically Weighted Regression*) quanto de Poisson (regressão global de Poisson, GWPR – *Geographically Weighted Poisson Regression* e GWPR *semi-paramétrico*). A análise exploratória revelou que não há evidência nem de autocorrelação espacial nem de heterogeneidade espacial da taxa de homicídios, por isso não se justificou a sua análise através de modelos de regressão espacial. Já os resultados da taxa de roubos, pelo contrário, apontam a existência de ambas características. Na sequência foram estimados diversos modelos de regressão para a taxa de roubo. Entre os modelos lineares, OLS e GWR, o modelo GWR obteve uma performance melhor. Porém, em ambos modelos, os resíduos apresentaram autocorrelação espacial, o que quebra a premissa de independência no caso do OLS e indicia um modelo GWR mal especificado. Foram então implementados modelos de regressão de Poisson, os quais são adequados à modelação de dados de contagem e taxas. O modelo GWPR apresentou um melhor ajustamento do que o modelo global de Poisson e do que os modelos GWPR semi-paramétricos. Os resultados indicam que os fatores população, renda, ocupação, educação e ambiente familiar são importantes para explicar as causas da taxa de roubo no Rio Grande do Sul, e que o grau de relevância de cada um deles varia espacialmente.

ABSTRACT

Brazil has high crime rates, with the homicide rate being one of the highest in the world. Despite remaining low compared to other states, Rio Grande do Sul has had a worrying growth in crime rates in recent years. This study aims to investigate Rio Grande do Sul's crime patterns, especially the robbery and homicide rates, as well as the processes that influence these patterns. For this purpose, different regression models were investigated, both linear (OLS – *Ordinary Least Squares*, and GWR – *Geographically Weighted Regression*) and Poisson models (global Poisson regression, GWPR – *Geographically Weighted Poisson Regression*, and *semiparametric GWPR*). Exploratory analysis revealed that there is no evidence of either spatial autocorrelation or spatial heterogeneity of the homicide rate, so their analysis through spatial regression models was not justified. The results of robbery rate, in contrast, point to the existence of both features. Following, several spatial regression models for robbery rate were estimated. Among the linear models, OLS and GWR, the GWR model had a better performance. However, the residuals presented spatial autocorrelation in both models, which contradicts the assumption of independence of OLS, and indicates a poorly specified GWR model. Poisson regression models were then implemented, as they are appropriate to model count and rate data. The GWPR model had a better fit than the global Poisson regression and also than the semiparametric GWPR models. The results indicate that population, income, occupation, education and family environment are important factors to explain the robbery rate in Rio Grande do Sul, and their degree of relevance varies in space.

PALAVRAS-CHAVE

Sistemas de Informação Geográfica

Taxa de roubo

Regressão Espacial

Regressão Geograficamente Ponderada

Regressão de Poisson

KEYWORDS

Geographical Information Systems

Robbery rate

Spatial Regression

Poisson Regression

Geographically Weighted Regression

ACRÓNIMOS

AdjR² – Coeficiente de determinação (R^2) ajustado

AICc – Critério de Informação de Akaike corrigido

GWR – Geographically Weighted Regression [Regressão Geograficamente Ponderada]

GWPR – Geographically Weighted Poisson Regression [Regressão de Poisson Geograficamente Ponderada]

IDH – Índice de desenvolvimento humano

IDHM – Índice de desenvolvimento humano municipal

IBGE – Instituto Brasileiro de Geografia e Estatística

IPEA – Instituto de Pesquisa Econômica Aplicada

OLS – Ordinary Least Squares [Mínimos Quadrados Ordinários]

SIG – Sistemas de Informação Geográfica

RS – Rio Grande do Sul

S-GWPR – Semiparametric Geographically Weighted Poisson Regression [Regressão de Poisson Geograficamente Ponderada semi-paramétrica]

ÍNDICE DO TEXTO

1. INTRODUÇÃO	1
1.1 Contexto do problema	1
1.2 Objetivos	2
1.3 Abordagem metodológica	2
1.4 Estrutura do documento	3
2. REVISÃO DE LITERATURA.....	4
2.1 Modelação da criminalidade	5
2.2 A criminalidade no Brasil	10
3. REGIÃO DE ESTUDO.....	26
4. MATERIAIS E MÉTODOS	29
4.1 Recolha e o pré-processamento dos dados	30
4.2 Análise exploratória	35
4.3 Modelos OLS	36
4.4 Modelo GWR.....	39
4.5 Modelos de Poisson	40
5. RESULTADOS	43
5.1 Análise exploratória da criminalidade associada a roubos.....	43
5.1.1 Estatísticas descritivas	44
5.1.2 Autocorrelação espacial	46
5.1.3 Hot Spots	48
5.2 Análise exploratória da criminalidade associada a homicídios.....	49
5.2.1 Estatísticas descritivas	50
5.2.2 Autocorrelação espacial	52
5.2.3 Hot Spots	54
5.3 Relação entre a criminalidade e potenciais variáveis explicativas	55
5.4 Modelação da criminalidade associada a roubos	66
5.4.1 Modelos OLS	66
5.4.2 Modelo GWR.....	70
5.4.3 Modelos Poisson	72
6. CONCLUSÃO.....	84
6.1 Limitações do estudo	85
6.2 Trabalhos futuros.....	86
REFERÊNCIAS	87
ANEXO	92

ÍNDICE DE TABELAS

Tabela 1: Resumo de alguns estudos relevantes sobre a criminalidade no Brasil.....	15
Tabela 2 - Variáveis independentes utilizadas em modelos de criminalidade no Brasil.	22
Tabela 3 - Variáveis utilizadas: abreviatura, descrição, ano e fonte.....	31
Tabela 4 - Legenda da Tabela 3: Grupos das variáveis independentes.....	35
Tabela 5 - Medidas descritivas do número anual de roubos de 2013 a 2017 no Rio Grande do Sul.	45
Tabela 6 - Medidas descritivas da taxa anual de roubos por 100 mil habitantes de 2013 a 2017 no Rio Grande do Sul.	45
Tabela 7 - Box plot do número de roubos por 100 mil habitantes de 2013 a 2017 no Rio Grande do Sul.	46
Tabela 8 - Medidas descritivas do número anual de homicídios no Rio Grande do Sul.	51
Tabela 9 - Medidas descritivas da taxa de homicídios por 100 mil habitantes no Rio Grande do Sul.	51
Tabela 10 - Matriz de correlação linear entre o homicídio, o roubo, a taxa de homicídio, a taxa de roubo e a população para os anos de 2013 a 2017 no estado.....	57
Tabela 11 - Matriz de correlação linear entre todas variáveis independentes e as variáveis dependentes (em negrito).....	58
Tabela 12 - Resultados e diagnósticos do modelo GWR.	70

ÍNDICE DE FIGURAS

Figura 1 - Mapa de contextualização do Rio Grande do Sul e sua capital Porto Alegre.....	26
Figura 2 - Taxa de Homicídios nos estados brasileiros.....	27
Figura 3 - Mapa da distribuição da taxa de roubo no estado em 2013.	44
Figura 4 - Relatório do teste de Global Moran's I para taxa de roubos por 100 mil habitantes em 2013 no estado.	47
Figura 5 - Mapa do teste de Local Moran's I da taxa de roubo no estado em 2013 (os clusters e outliers espaciais indicam autocorrelação espacial local significativa ao nível de significância de 5%).	48
Figura 6 - Mapa de Hot Spots da taxa de roubo no estado.....	49
Figura 7 - Mapa da distribuição da taxa de homicídio no estado em 2013.	50
Figura 8 - Boxplot da taxa anual de homicídios por 100 mil habitantes.	52
Figura 9 - Relatório do teste Global Moran's I para taxa de homicídios por 100 mil habitantes.	53
Figura 10 - Mapa do teste de Local Moran's I da taxa de homicídio no estado em 2013 (os clusters e outliers espaciais indicam autocorrelação espacial local significativa ao nível de significância de 5%).	54
Figura 11 - Mapa de Hot Spots da taxa de homicídios no estado em 2013.	55
Figura 12 - Gráfico de dispersão do número absoluto de homicídios, número de roubos e população em 1 milhão para os anos de 2013 a 2017.....	56
Figura 13 - Gráfico de dispersão da taxa de homicídio e de roubo por 100 mil habitantes e da população em 1 milhão, para os anos de 2013 a 2017.	56
Figura 14 - Resultados do modelo OLS.	66
Figura 15 - Resultados e diagnósticos do modelo OLS.....	67
Figura 16 - Relatório do teste de Moran Global I sobre os resíduos padronizados do modelo OLS.....	68
Figura 17 - Relatório do teste de Moran Global I sobre os resíduos padronizados do modelo GWR.	71
Figura 18 - Resultado Poisson Global.....	72
Figura 19 - Resultado Poisson Local (GWPR).....	73
Figura 20 - Resultado da ferramenta Local-To-Global.	73
Figura 21 - Mapa do coeficiente da variável população estimada por município no RS.....	74
Figura 22 - Mapa do coeficiente da variável PREN20RICO por município no RS.	75
Figura 23 - Mapa do coeficiente da variável RENOCUP por município no RS.	76
Figura 24 - Mapa do coeficiente da variável T_ATRASO_0_MED por município no RS.	77
Figura 25 - Mapa do coeficiente da variável T_DENS por município no RS.....	78
Figura 26 - Mapa do coeficiente da variável T_DES18M por município no RS.	79
Figura 27 - Mapa do coeficiente da variável T_FUNDIN_TODOS por município no RS.	80
Figura 28 - Mapa do coeficiente da variável T_MULCHEFEFIF014 por município no RS.....	81
Figura 29 - Mapa do coeficiente da variável T_NESTUDA_NTRAB_MMEIO por município no RS.	82
Figura 30- Mapa do percentual de <i>deviance</i> explicada pelo modelo por município no RS.....	83
Figura EE - Gráfico de dispersão entre a taxa de roubo e a população estimada.	92
Figura FF - Gráfico de dispersão entre a taxa de roubo e a TAXAHOMEM15A24POP.	92
Figura GG - Gráfico de dispersão entre a taxa de roubo e a T_M15A17CF.	93
Figura HH - Gráfico de dispersão entre a taxa de roubo e o IDHM.	93
Figura II - Gráfico de dispersão entre a taxa de roubo e a T_OCUPDESLOC_1.....	93
Figura JJ - Gráfico de dispersão entre a taxa de roubo e a T_FUNDIN_TODOS.	94
Figura KK - Gráfico de dispersão entre a taxa de roubo e a T_MULCHEFEFIF014.	94
Figura LL - Gráfico de dispersão entre a taxa de roubo e a T_DENS.....	94
Figura MM - Gráfico de dispersão entre a taxa de roubo e a T_NESTUDA_NTRAB_MMEIO.	95
Figura NN - Gráfico de dispersão entre a taxa de roubo e a T_DES18M.....	95
Figura OO - Gráfico de dispersão entre a taxa de roubo e a PIND.	95
Figura PP - Gráfico de dispersão entre a taxa de roubo o RENOCUP.....	96
Figura QQ - Gráfico de dispersão entre a taxa de roubo o GINI.	96
Figura RR - Gráfico de dispersão entre a taxa de roubo e o PREN20RICOS.	96

Figura SS - Gráfico de dispersão entre a taxa de roubo e a T_ATRASO_0_MED.	97
Figura TT - Gráfico de dispersão entre a taxa de roubo e a T_FREQ18A24.....	97
Figura UU - Gráfico de dispersão entre a taxa de roubo e a T_FREQ15A17.....	97
Figura VV - Gráfico de dispersão entre a taxa de roubo e a T_FLMED.	98
Figura WW - Gráfico de dispersão entre a taxa de roubo e a T_ANALF15M.	98
Figura XX - Mapa de <i>std residual</i> do modelo GWPR por município no RS.	99
Figura YY - Mapa dos erros padrão para a variável população estimada no modelo GWPR.	99
Figura ZZ - Mapa dos erros padrão para a variável PREN20RICO no modelo GWPR.	100
Figura AAA - Mapa dos erros padrão para a variável RENOCUP no modelo GWPR.	100
Figura BBB - Mapa dos erros padrão para a variável T_ATRASO_0_MED no modelo GWPR.....	101
Figura CCC - Mapa dos erros padrão para a variável T_DENS no modelo GWPR.	101
Figura DDD - Mapa dos erros padrão para a variável T_DES18M no modelo GWPR.	102
Figura EEE - Mapa dos erros padrão para a variável T_FUNDIN_TODOS no modelo GWPR.....	102
Figura FFF - Mapa dos erros padrão para a variável T_MULCHEFEFIF014 no modelo GWPR.	103
Figura GGG - Mapa dos erros padrão para a variável T_NESTUDA_NTRAB_MMEIO no modelo GWPR.	103

1. INTRODUÇÃO

1.1 Contexto do problema

O debate sobre a segurança pública no Brasil tem sido recorrente. O país possui altas taxas de criminalidade, sendo a taxa de homicídios no Brasil uma das mais altas do mundo, e esta realidade vem piorando. Em 2016 os homicídios no Brasil atingiram uma marca histórica com uma taxa de 30,3 homicídios para cada 100 mil habitantes, cerca de 30 vezes a taxa da Europa (Cerqueira et al., 2018, IPEA). No estado do Rio Grande do Sul não é diferente, apesar de permanecer com valores baixos em relação a outros estados, o estado teve um crescimento abrupto nas taxas de criminalidade nos últimos anos. A taxa de homicídios para cada 100 mil habitantes em 2006 no estado foi de 18,1, já em 2016 essa taxa bateu sua máxima de 28,6 (Cerqueira et al., 2018, IPEA).

A criminalidade tem um forte impacto na sociedade e, além dos prejuízos econômicos gerados na forma de gastos públicos e privados com a segurança, retração do turismo, dos investimentos e do valor dos imóveis nas localidades, também afeta o bem-estar social. O sentimento de insegurança e medo levam a mudanças de comportamento e estilo de vida, a preocupação é constante (W. Batella, 2008; Moreira & Fochezatto, 2017; Pinto & Coronel, 2015; Saraiva, Conceição, & França, 2017). Segundo o IPEA (Instituto de Pesquisa Econômica Aplicada) (2019), que se dedica a 20 anos ao estudo da violência e tem realizado anualmente o "Atlas da violência", a violência constitui uma das maiores questões de políticas públicas no Brasil.

Sendo assim, o combate à criminalidade se tornou um dos principais temas discutidos no país, sendo amplamente abordado nos meios de comunicação e em campanhas eleitorais. Mas para se propor novas políticas e estratégias de redução à criminalidade é necessário compreender suas causas e padrões. É aqui que este trabalho se insere, tendo como objetivo principal investigar os padrões da criminalidade no Rio Grande do Sul e as possíveis causas para os padrões observados.

1.2 Objetivos

Os principais objetivos do trabalho são descrever os padrões da criminalidade no Rio Grande do Sul, em particular das taxas de roubo e de homicídio, e investigar possíveis causas para os padrões observados em 2013. Pretende-se também descrever a evolução destes tipos de crime entre 2013 e 2017.

As principais questões que o trabalho se propõe a investigar se encontram listadas abaixo:

- A criminalidade possui autocorrelação espacial?
- Quais são os fatores socioeconômicos que afetam a criminalidade?
- A relação entre os fatores socioeconômicos e a criminalidade varia espacialmente?

Para atingir os objetivos propostos, serão construídos e comparados modelos de regressão linear e de Poisson, tanto globais como locais (i.e. espaciais). Serão identificadas potenciais variáveis explicativas através de uma revisão da literatura sobre o tema em estudo e dos dados disponíveis, assumindo que o padrão dessas variáveis não sofreu grandes alterações entre 2010 e 2013.

1.3 Abordagem metodológica

Embora este trabalho não seja pioneiro na utilização de técnicas de análise espacial e geoestatística para entender a criminalidade no Rio Grande do Sul, assume-se aqui não somente que a criminalidade varia no espaço, mas também os seus processos causais, ou seja, as relações entre as variáveis explicativas e a variável dependente podem ser não-estacionárias.

Portanto, o enfoque é explorar as variáveis preditivas, com embasamento teórico, e analisá-las com diferentes tipos de modelos de regressão. Nomeadamente, o modelo global estimado por Ordinary Least Squares (OLS, em português os Mínimos Quadrados Ordinários), que considera as variáveis estacionárias e a independência das observações, o modelo Geographically Weighted Regression (GWR, em português

Regressão Geograficamente Ponderada), que leva em conta a não-estacionariedade, e por último o modelo de regressão de Poisson local (GWPR), que também considera a não-estacionariedade, mas leva em conta dados de contagem ou taxa.

Na literatura, os modelos de regressão espacial local que consideram a não-estacionariedade têm mostrado maior potencial explicativo, portanto têm mais relevância para aplicações de políticas no mundo real. Na verdade, se existem variações nos processos causais pelo território e estas variações não são levadas em conta, então o modelo estatístico será impreciso (Cahill & Mulligan, 2007). Este trabalho tem como propósito contribuir com a investigação de variáveis explicativas do crime através desta perspectiva e, assim, também contribuir para a validação da abordagem com regressões espaciais locais. Para tal, se propõe a comparar os diferentes tipos de modelo de regressão para investigar qual destes tem uma melhor performance e se ajusta melhor nos dados de criminalidade.

1.4 Estrutura do documento

No que diz respeito a estrutura do trabalho, este começa com uma breve revisão da literatura sobre modelação da criminalidade e sobre a criminalidade no Brasil, no capítulo 2, onde são apresentadas as principais metodologias utilizadas para modelar a criminalidade, as variáveis e os principais resultados. Seguido do capítulo 3, uma breve descrição da região de estudo. No capítulo 4, são detalhados os materiais e métodos. A seguir, no capítulo 5, são apresentados e discutidos os resultados da investigação dos padrões do crime, incluindo a modelação da criminalidade associada a roubos através de modelos de regressão linear (OLS e GWR) e de Poisson (global e GWPR). Por fim, no capítulo 6, são apresentadas as principais conclusões, bem como as limitações do estudo e sugestões para trabalhos futuros.

2. REVISÃO DE LITERATURA

A abordagem espacial em estudos de criminalidade não é novidade, ainda que continue atual. Segundo Bernasco & Elffers (2010) remonta ao século XIX quando se comprova que a criminalidade varia de acordo com o lugar, utilizando-se mapas para visualizar estas diferenças e, ainda, explorando-se as relações estatísticas entre crime, pobreza e educação. Esta área foi impulsionada por volta das décadas de 1970 e 1980 pelo desenvolvimento das tecnologias de Sistemas de Informação Geográfica (SIG), pela Estatística Multivariada, por estudos ecológicos, que evidenciam o papel do ambiente na criminalidade, e pela teoria econômica do crime. Nos Estados Unidos, na década de 1970, surge uma subdisciplina da Geografia chamada Geografia do Crime (W. Batella, 2008).

Quanto ao desenvolvimento deste domínio, Batella (2008) tece que a análise espacial, consolidada na década de 1950, foi uma grande contribuição, proporcionando reflexões focadas nos problemas de proximidade, concentração e dispersão da incidência criminal e na investigação dos padrões de distribuição. A ideia de que os crimes não ocorrem aleatoriamente no tempo e no espaço foi revolucionária e mudou o enfoque de investigação, antes dos criminosos, para o delito propriamente dito.

Estes avanços tanto tecnológicos quanto teóricos contribuíram para a construção de metodologias sólidas que permitem análises mais complexas e aprofundadas da criminalidade. O entendimento de que o ambiente influencia a criminalidade e o desenvolvimento de técnicas para se estimar o quanto, levou a uma série de pesquisas nas últimas décadas que procuram explicar as causas e padrões da criminalidade através de modelos de regressão com variáveis socioeconômicas e testes estatísticos para investigar a presença de autocorrelação espacial. Os resultados têm demonstrado que criminalidade é fortemente influenciada por condições do contexto (ambiente) e que o crime possui autocorrelação espacial, ou seja, os crimes não ocorrem aleatoriamente no espaço.

2.1 Modelação da criminalidade

Construir uma estrutura teórica sólida para explicar a criminalidade e suas causas é uma tarefa complexa, pois existem várias correntes teóricas (W. Batella, 2008; Oliveira, 2008). Porém, destacam-se a teoria econômica de Becker (1968) com uma grande contribuição para o entendimento das motivações do crime, e a abordagem ecológica proposta por Bronfenbrenner (1996), que, embora não tenha surgido sob a temática da criminalidade, é muito utilizada neste âmbito. Em seu trabalho, Oliveira (Oliveira, 2008) aponta que para uma análise mais completa sobre o que, de fato, leva indivíduos a cometerem atos ilícitos, a abordagem ecológica deve complementar a teoria econômica, uma vez que esta pode ter um caráter excessivamente financista.

Na teoria econômica, o crime seria exercido racionalmente em função da vantagem econômica para o indivíduo que o pratica. Os indivíduos procuram otimizar seu nível de bem-estar através de atividades ilegais, isso ocorre principalmente se as atividades lícitas ao seu alcance não oferecem um retorno melhor ou suficiente (Moreira & Fochezatto, 2017; Oliveira, 2008; Santos & Kassouf, 2008). Os benefícios do crime têm de ser maiores do que os seus custos, são estes: os custos de oportunidade, custo de execução e planejamento e a probabilidade da punição (Oliveira, 2008). Segundo Santos & Kassouf (2008, p. 346) a teoria econômica sustenta que *"alguns indivíduos tornam-se criminosos não porque suas motivações básicas são diferentes das de outros indivíduos, mas porque seus custos e benefícios diferem"*.

A abordagem ecológica tem como ideia central que o ambiente (contexto) possui um importante papel no desenvolvimento do indivíduo e quando aplicado à criminalidade significa que o processo entre a inserção do indivíduo no ambiente e seu histórico pessoal é determinante para a decisão de praticar um ato criminoso (Oliveira, 2008). De acordo com Cahill & Mulligan (2007) a compreensão das causas da criminalidade seria incompleta sem a perspectiva dos estudos ecológicos, que levam em conta as características, físicas e sociais, do ambiente. Nesta concepção fatores individuais e a estrutura social desempenham um papel importante na criminalidade.

Neste sentido, Oliveira (2008) propõe a adição, dentro da teoria econômica, de um custo moral e explica que este é determinado pelo desenvolvimento do indivíduo e seu processo de construção do julgamento moral. Neste processo os valores são construídos a partir da história e do ambiente que cerca o indivíduo, segundo a abordagem ecológica. O custo moral seria uma barreira para o crime, pois, mesmo que haja muitos benefícios, se o indivíduo possui um custo moral maior, via de regra, ele não irá praticar o ato criminoso (Oliveira, 2008). A estrutura teórica do presente trabalho se baseia nesta combinação entre a teoria econômica e a abordagem ecológica.

Há também outras correntes teóricas, exemplo as que focam mais na vizinhança do que no indivíduo, como a teoria da desorganização social, que enfatiza que a criminalidade seria uma consequência da falta de capacidade da comunidade para gerar um sistema efetivo de controle institucional (Cahill & Mulligan, 2007; Lee, Kang, & Kim, 2009). Ou, também, a teoria da oportunidade, na qual se acredita que uma área terá maior criminalidade conforme a presença de oportunidades, em uma combinação no tempo e espaço de alvos adequados, infratores motivados e a falta de proteção (guarda). Nessas teorias, as motivações e causas do crime não importam, e sim o quanto uma área é propícia a ter atividades criminosas e o quanto a comunidade é preparada (capacitada) para lidar com o crime. Acredita-se que essas teorias não necessariamente se excluem, porém, o foco deste trabalho é nas causas da criminalidade.

Segundo Santos & Kassouf (2008, p. 350) "*diversas variáveis socioeconômicas têm sido testadas na investigação empírica do crime, entre elas: renda, taxa de desemprego, nível de escolaridade, pobreza, desigualdade de renda e urbanização*", porém não se tem encontrado um consenso sobre a influência destas e, por vezes, o resultado da direção (sinal) de variáveis é contrário ao esperado. Isto pode se dar pelo fato de algumas variáveis serem ambíguas, por exemplo, uma maior renda *per capita* pode sugerir que a região tenha menor taxa de criminalidade, pois sugere que quem reside ali tem maior retorno no mercado lícito, porém pode representar o contrário, pois também se associa a um maior número de vítimas potenciais.

Modelar a criminalidade, portanto, não é uma tarefa simples, pois envolve uma infinidade de fatores, e alguns são, ainda por cima, muito difíceis de mensurar. Exemplos de fatores que, também, podem influenciar (negativamente) a criminalidade são: a liberação da posse de armas; o aumento do tráfico de drogas; e a morosidade (ou ausência) da justiça, que leva a grande rotatividade dos criminosos entre a prisão e a liberdade, frequentemente liberados sem pagar pelos seus crimes (Lima, Ximenes, Ramos de Souza, Luna, & Albuquerque, 2005; Moreira & Fochezatto, 2017). Este último está diretamente ligado com a probabilidade de punição, porém é especialmente difícil de se quantificar e, portanto, de levar em consideração no modelo. Lima et al (2005, p. 177) ressalta que *"mesmo no caso de um único modelo vir a abarcar toda essa complexidade, não poderia ser generalizado para além da situação particular alvo da análise"*.

No entanto, tentar modelar a criminalidade é importante para se compreender melhor o fenômeno e suas causas principais e, assim, elaborar planos de ação mais direcionados e eficazes. Na prática, as técnicas geralmente utilizadas para isto são testes estatísticos que investigam a presença de autocorrelação espacial, sobretudo com as estatísticas de Moran Global e Local, análise de Hot Spots, e modelos de regressão linear múltipla (que designaremos por modelo OLS), embora atualmente a tendência é a utilização de modelos que têm em consideração a não-estacionariedade das variáveis, como o modelo GWR (e.g., Cahill & Mulligan, 2007; Lee et al., 2009; Rybarczyk, Maguffee, & Kruger, 2015; Vilalta & Muggah, 2016). Estes autores salientam o aumento do poder explicativo dos modelos GWR ao incorporar relações espaciais locais.

Cahill & Mulligan (2007) realizaram um modelo global e um modelo local para estudar a violência em Portland, Oregon, e compararam-nos. A violência foi considerada em forma de taxa agregada por bloco de censo, incluindo homicídio, agressão sexual, roubo e agressão agravada (*aggravated assault*), para a média dos anos de 1998 a 2002. Para ambos os modelos as variáveis independentes foram o índice de heterogeneidade, famílias monoparentais (single-person household), índice de concentração nos extremos (calculado como o número de famílias ricas menos o número de famílias pobres), famílias casadas, uso múltiplo de terra, estabilidade

residencial (percentagem de residentes que moravam na mesma residência em 1995) e densidade populacional. Seus resultados demonstram que o modelo GWR pode ajudar a explicar os resultados do modelo OLS, em particular resultados contra intuitivos no modelo global, pois este pode mascarar importantes variações nos parâmetros. Os resultados do modelo GWR revelam que a relação entre a taxa de violência e metade das variáveis independentes variam no espaço.

Lee et al (2009) exploram os determinantes do crime em Seoul, Coréia do Sul. O crime foi considerado como um agregado de cinco tipos principais de crime: homicídio, roubos, estupro, agressão e furto; juntamente com crime intelectual. As variáveis explicativas foram subdivididas em variáveis socioeconômicas (imposto sobre a propriedade por agregado familiar, população estabelecida, população flutuante, proporção de jovens de 15 a 24 anos, proporção de graduados em universidade e número de policiais) e variáveis de planejamento urbano (proporção de área de desenvolvimento limitada, proporção de área residencial e proporção de área de alojamento). Os autores concluem que o crime é um fenômeno regional, discernindo o efeito local e global de cada variável sobre a incidência de crime, e afirmam que características regionais afetam claramente a incidência de crimes.

Rybarczyk et al (2015) investigaram a influência da saúde pública (média do índice de massa corporal), condições socioeconômicas (número de casas ocupadas pelo locatário, número de famílias que falam inglês, número de residentes não-brancos, número de famílias abaixo do nível de pobreza e moradores com um diploma do ensino médio), capital social (atitude em relação à própria saúde), estresse ambiental (níveis de chumbo no sangue) e contexto de vizinhança (comprimento de trilhas de bicicleta, lotes vagos, comprimento das ferrovias, comprimento das calçadas e pontos de venda de alimento) em Flint, Michigan, através dos modelos global OLS e local GWR. A taxa de crime considera agressão agravada, homicídio, roubos e todas as forma de conduta sexual criminosa, e é normalizada por setor censitário. Suas conclusões também corroboram a importância dos modelos locais, que revelam nuances das relações entre o crime e as variáveis explicativas, e ainda o modelo GWR obteve uma performance melhor que o modelo OLS.

Vilalta & Muggah (2016) modelaram espacialmente a criminalidade na área metropolitana da Cidade do México através dos métodos OLS e GWR. A criminalidade considerada foi a taxa de investigações criminais iniciadas em 2010, os tipos de crimes não foram especificados. As variáveis independentes foram escolhidas como os proxies da teoria da desorganização social (índice de desigualdade social, porcentagem de habitantes que viviam em outro estado nos últimos 5 anos, bares e restaurantes por 100 mil habitantes e porcentagem de agregados familiares chefiados por mulheres) e da teoria da anomia¹ (porcentagem de participação eleitoral no Congresso Federal, índice de Gini, retenção de grau de igualdade social na escola primária e porcentagem de agregados familiares chefiados por mulheres). Suas conclusões mostram que os municípios da área metropolitana da Cidade do México não são homogêneos quanto às taxas de criminalidade nem suas causas (e, portanto, soluções). Consequentemente, um modelo global não seria adequado para representar essa realidade e sim um modelo local, que permite analisar as especificidades locais, melhor determinando as causas e soluções da criminalidade de uma determinada região.

Outro modelo de regressão que permite ter em conta a não-estacionariedade das relações espaciais é o modelo GWPR – Geographically Weighted Poisson Regression (em português, Regressão de Poisson Geograficamente Ponderada). Este modelo tem sido utilizado na literatura sobre modelação da criminalidade (Chen et al., 2017; Vilalta & Fondevila, 2019; Vilalta, Sanchez, Fondevila, & Ramirez, 2019). Os modelos de regressão Poisson são frequentemente usados para modelar dados de contagem e taxas, porque se baseiam na distribuição de Poisson. Osgood (2000) publicou "*Poisson-Based Regression Analysis of Aggregate Crime Rates*" onde sugere o uso de regressão de Poisson para lidar com os problemas ligados a taxas de crime e baixa população em unidades agregadas (como bairros, cidades, etc.).

¹ A anomia refere-se à ausência ou desintegração das normas sociais. Segundo a teoria da anomia de Robert K. Merton, o crime caracteriza uma situação de anomia social.

2.2 A criminalidade no Brasil

No âmbito da revisão bibliográfica, foi efetuada uma pesquisa para investigar o que há de literatura sobre criminalidade no Brasil, com enfoque espacial. O intuito foi examinar quais metodologias são geralmente empregadas e quais foram os principais resultados, e não esgotar a literatura existente. Após a pesquisa, se constatou que as publicações sobre criminalidade no Brasil que envolvem análise espacial são bastante atuais, aparecendo em meados dos anos 2000 (Bartz, Quartieri, & Menezes, 2018; W. Batella, 2008; W. B. Batella & Diniz, 2010; Carrets, Oliveira, & Menezes, 2018; Gomes, Evangelista, Lima, & Parré, 2017; Lima et al., 2005; Moreira & Fochezatto, 2017; Nery, Peres, Cardia, Vicentin, & Adorno, 2012; Oliveira, 2008; Saraiva et al., 2017; Sass, Porsse, & Silva, 2016).

Nestes trabalhos, os autores convergem sobre o aumento preocupante da criminalidade no Brasil e procuram investigar seus padrões e causas, afim de auxiliar medidas e políticas públicas para reduzi-la. Destaca-se o trabalho de Oliveira (2008) que propõe uma forte estrutura teórica para as causas da criminalidade e é citado em muitas das publicações brasileiras sobre este tema. A seguir são feitas algumas constatações sobre as metodologias utilizadas nestes trabalhos.

Uma parte importante deste tipo de trabalho é a escolha de como representar a criminalidade. Com frequência, os autores utilizam a variável homicídio para essa finalidade. Alguns autores, como Saraiva et al (2017) e Moreira e Fochezatto (2017), usam um agregado de tipos diferentes de crime para representar a criminalidade de forma geral. Outra forma, muito comum, é considerar diferentes tipos de crime separadamente, como, por exemplo: homicídio, roubo e furto. Essa abordagem permite uma análise mais detalhada de cada modalidade de crime.

Para área de estudo, no Brasil, é muito usual a utilização de um estado, comparando os valores de criminalidade para cada município. Oliveira (2008) acredita que esta realidade é, provavelmente, uma consequência da maior disponibilidade de dados sobre crime a partir das secretarias estaduais. Porém, Nery et al (2012) atentam sobre a importância de usar dados mais desagregados possível para melhor caracterizar a heterogeneidade local e evitar conclusões simplistas. Em seu trabalho

utilizam o município como área de estudo, comparando a criminalidade nos setores censitários, menor unidade agregada de território no Brasil.

As metodologias geralmente empregadas incluem a visualização da distribuição da criminalidade na área de estudo, a investigação da ocorrência de autocorrelação espacial através do teste estatístico de Moran Global e a verificação dos padrões locais através de técnicas LISA (Local Indicators of Spatial Association), em particular a estatística de Moran Local (Anselin, 1995). Além disso, a maior parte das publicações também investiga os fatores determinantes da criminalidade através da correlação linear entre as variáveis e do emprego de modelos de regressão, sobretudo através do método OLS. Ressalta-se os trabalhos de Sass et al (2016) e Gomes et al (2017) que utilizam o modelo de regressão GWR. Embora poucos trabalhos sobre criminalidade no Brasil considerem a não-estacionariedade das variáveis explicativas, a tendência é que cada vez haja mais investigação nesta direção, pois os resultados têm sido reveladores.

Sass et al (2016) procuraram identificar os fatores que influenciam as taxas de homicídios dos municípios do estado do Paraná através de modelos globais e o modelo local GWR. Os fatores considerados no modelo foram percentagem de homens entre 15 e 24 anos na população, densidade demográfica, grau de urbanização, índice de Gini, taxa de analfabetismo, taxa de desemprego, taxa de pobreza e taxa de policiais por 100 mil habitantes, este último relacionado com o efeito de dissuasão e atuação do poder público. A taxa de pobreza, desemprego e o índice de Gini teriam a função de captar o efeito esperado pela teoria da desorganização social. Os resultados mostram que o estado do Paraná apresenta autocorrelação espacial para criminalidade. As variáveis grau de urbanização e taxa de pobreza, dentre as outras variáveis explicativas, apresentaram maior impacto nas taxas de homicídio. Por fim concluem que o modelo local GWR apresenta melhores resultados em relação aos modelos globais, isso porque o impacto (influência) das variáveis explicativas, como por exemplo a taxa de pobreza, não é uniforme pelo território. Portanto o modelo GWR fornece informações mais relevantes e detalhadas para elaboração de políticas públicas mais efetivas.

Gomes et al (2017) exploraram os determinantes do crime nos municípios de Minas Gerais, para o ano de 2010, utilizando o modelo GWR. Os autores subdividiram a criminalidade em quatro variáveis de acordo com o tipo de crime: taxa de crimes violentos (homicídio, homicídio tentado, estupro, roubo e roubo à mão armada), taxa de crimes violentos contra o patrimônio (roubo e roubo à mão armada), taxa de crimes violentos contra a pessoa (homicídio, homicídio tentado e estupro) e taxa de homicídios. As variáveis explicativas escolhidas foram: a existência de estabelecimento penitenciário ou prisional, gasto per capita com infraestrutura, gasto per capita com segurança pública, gasto per capita com desenvolvimento econômico, habitantes por policial civil ou militar, população de 15 a 24 anos, população de 25 a 29 anos, renda per capita, taxa de frequência ao ensino médio, densidade demográfica e taxa de armas apreendidas. Os resultados revelam que os crimes contra a pessoa tendem a ser mais frequentes em áreas economicamente menos desenvolvidas, ao contrário dos crimes contra o patrimônio, que tendem a ocorrer mais em regiões ricas, e a maior concentração de população entre 15 e 24 anos tende a aumentar a criminalidade de maneira geral. Concluem também que o modelo GWR é mais interessante, pois seus resultados, baseados em valores individuais, tendem a dar subsídios relevantes para a elaboração de medidas e políticas mais eficazes ao combate da criminalidade.

Os resultados destes trabalhos têm revelado, de forma geral, a presença de autocorrelação espacial para a criminalidade, corroborando com a lei de Tobler: "tudo está relacionado, porém coisas próximas estão mais relacionadas do que coisas distantes" e sugerem haver um efeito de transbordamento (Oliveira, 2008; Saraiva et al., 2017). Oliveira (2008, p. 38) afirma que o "[...] crime é algo que se difunde e que a criminalidade de uma cidade pode afetar uma cidade vizinha". Portanto, essa dependência espacial deve ser levada em conta quando se modela a criminalidade. Mais atualmente, têm se verificado que a relação entre as variáveis explicativas e a variável dependente também variam no espaço e, portanto, modelos espaciais locais seriam mais indicados.

Quanto aos determinantes do crime no Brasil, as variáveis frequentemente utilizadas são relacionadas à educação, renda, desenvolvimento humano,

desigualdade de renda, desorganização familiar e condições de moradia. Como já referido anteriormente, os resultados sobre a influência de variáveis socioeconômicas na criminalidade não são consistentes entre si (Lima et al., 2005; Santos & Kassouf, 2008). Isso pode ocorrer, pois a criminalidade tem causas diferentes em locais diferentes e também pode ser resultado de como cada variável é representada.

Um exemplo disto é a variável vínculo empregatício utilizada como variável independente no trabalho de Moreira & Fochezatto (2017), onde não teve significância no modelo, e em Saraiva et al (2017), onde teve significância no modelo. Porém, é importante ressaltar que os indicadores eram diferentes, Moreira & Fochezatto (2017) utilizaram o vínculo per capita, total de indivíduos empregados no mercado formal de trabalho dividido pela população total. Já Saraiva et al (2017) utilizaram o vínculo empregatício até 3 meses, percentual de vínculos com até três meses de duração sobre o total de vínculos, onde a ênfase é na duração dos vínculos empregatícios (rotatividade). A maior rotatividade, curta duração dos vínculos empregatícios, se mostrou importante entre as causas da criminalidade, ao contrário do vínculo per capita apenas. Portanto, os indicadores utilizados para representar cada variável podem afetar o resultado final.

Se constatou, ao analisar os principais resultados dos estudos de criminalidade, que as variáveis socioeconômicas atuam, por vezes, diferentemente para diferentes tipos de crimes. Como, por exemplo, em Oliveira (2008) a variável habitações subnormais demonstra uma relação positiva significativa com roubos, porém a relação com furtos e homicídios é não significativa, ou a variável renda dos mais ricos que demonstra uma relação positiva significativa com roubos e furtos, porém não significativa com homicídios. Há, no entanto, algumas variáveis que têm se mostrado, via de regra, significativas para a criminalidade em geral, como a densidade demográfica, a desigualdade de renda, a frequência escolar e a taxa de urbanização.

Ressalta-se que os autores Lima et al (2005) e Moreira & Fochezatto (2017) atribuem, também, o aumento das taxas de criminalidade no Brasil ao sucateamento das instituições de segurança pública, a morosidade e, por vezes, ausência da justiça brasileira. Porém, como dito anteriormente, essas variáveis são difíceis de se adicionar

na modelação da criminalidade. Além disto, essas variáveis estão mais ligadas com a penalidade e o controle social, do que nas causas propriamente.

A Tabela 1 apresenta um resumo de alguns estudos mencionados anteriormente sobre a criminalidade no Brasil, salientando a metodologia, variáveis utilizadas e principais conclusões. A Tabela 2 coloca em evidência as variáveis independentes frequentemente utilizadas em modelos de criminalidade no Brasil, bem como a sua significância para diferentes tipos de crime.

Tabela 1: Resumo de alguns estudos relevantes sobre a criminalidade no Brasil.

Referência	Metodologia	Área de estudo	Variável Dependente (crime)	Variáveis Independentes	Resultados principais
Lima et al (2005)	Global Moran's Correlação Linear Conditional Auto Regressive (CAR)	Municípios do estado de Pernambuco	Taxa de homicídio para homens de 15 a 49 anos por 100 mil habitantes	Índice de Condições de Vida (ICV) Índice de desenvolvimento humano municipal (IDHM) Coeficiente de GINI Índice desigualdade de Theil Renda familiar per capita Renda média do chefe de família Índice de pobreza (percentual de pessoas de 10 anos ou mais de idade, por classes de rendimento nominal médio mensal abaixo de um salário mínimo) Taxa de analfabetismo (percentual de pessoas com 15 anos ou mais que não são capazes de ler ou escrever) Densidade demográfica	Presença de autocorrelação. Após analisar a matriz de correlação, as variáveis independentes que sobraram foram: taxa de analfabetismo e índice de pobreza. O modelo explicou 24,6% da variabilidade da variável dependente.

Referência	Metodologia	Área de estudo	Variável Dependente (crime)	Variáveis Independentes	Resultados principais
Batella (2008)	Distribuição espacial do crime Mapa topográfico do crime Global Moran's Local Moran's Correlação linear (Pearson)	Municípios do estado de Minas Gerais	Crimes violentos contra o patrimônio: roubo, roubo a mão armada, roubo seguido de morte, roubo de veículo e extorsão mediante sequestro. Crimes violentos contra a pessoa: homicídio consumado, homicídio tentado, estupro consumado, estupro tentado, sequestro e cárcere privado. Por 100 mil habitantes	Desenvolvimento humano: IDHM. Riqueza: Renda per capita, PIB per capita, bens de consumo TV, bens de consumo Telefone, bens de consumo Carro, bens de consumo Geladeira, bens de consumo Computador. Desigualdade de renda: Índice de GINI. Infra-estrutura: Percentual de pessoas que vivem em domicílios com água encanada, Percentual de pessoas que vivem em domicílios com energia elétrica, Percentual de pessoas que vivem em domicílios urbanos com serviço de coleta de lixo, taxa de urbanização. Educação: Taxa de alfabetização, taxa bruta de frequência à escola, percentual de pessoas de 15 anos ou mais analfabetas. Estrutura populacional: População total, população de 15 a 24 anos de idade, densidade demográfica. Imigração: Imigração Inter-estadual, imigração intra-estadual.	Presença de autocorrelação espacial para os dois tipos de crimes. Correlação positiva para crimes contra o patrimônio: desenvolvimento humano, riqueza, infra-estrutura, educação, estrutura populacional e imigração. Sem correlação para crimes contra o patrimônio: desigualdade de renda. Correlação positiva (baixa) para crimes contra a pessoa: riqueza, desigualdade de renda, estrutura populacional e imigração. Correlação negativa para crimes contra a pessoa: desenvolvimento humano, infra-estrutura, educação.

Referência	Metodologia	Área de estudo	Variável Dependente (crime)	Variáveis Independentes	Resultados principais
Oliveira (2008)	Distribuição espacial do crime Global Moran's LISA (Indicadores Locais de Associação Espacial) OLS LM (Multiplicador de Lagrange)	Municípios do estado do Rio Grande do Sul	Furto (crime contra o patrimônio) Roubo (crime contra o patrimônio e a pessoa) Homicídio (crime contra a pessoa)	Renda: renda média dos 10% mais ricos e renda média dos 20% mais pobres. Densidade demográfica. Mulheres chefes de família: percentual de famílias cuja única fonte de renda é originária de indivíduos do sexo feminino. Habitações subnormais: percentual de indivíduos que vivem em habitações em condições inapropriadas para moradia. Educação (não especificado). Desigualdade de renda.	Presença de autocorrelação para furtos e roubos e sem autocorrelação para homicídios. Para evitar multicolinearidade foram utilizados dois modelos, separando as variáveis escola e desigualdade de renda, pois estas possuem alta correlação com outras variáveis. Relação positiva e significativa com furtos: renda dos mais ricos, densidade demográfica, mulheres chefes de família, educação e desigualdade de renda. Relação positiva e significativa com roubos: renda dos mais ricos, densidade demográfica, mulheres chefes de família, habitações subnormais, educação e desigualdade de renda. Relação positiva e significativa com homicídios: densidade demográfica, mulheres chefes de família, educação e desigualdade de renda. Relação negativa e significativa com homicídios: renda dos mais pobres. O modelo para furtos explica 30% da variabilidade. O modelo para roubos explica 68% da variabilidade. O modelo para homicídios explica 7% da variabilidade.

Referência	Metodologia	Área de estudo	Variável Dependente (crime)	Variáveis Independentes	Resultados principais
Sass et al (2016)	Distribuição espacial do crime Global Moran's LISA OLS LM SAR (Spatial Autoregressive Model) SEM (Spatial Error Model) SMD (Modelo Dublin Espacial) GWR	Municípios do estado do Paraná	Taxa de homicídio por 100 mil habitantes.	Porcentagem de homens entre 15 e 24 anos na população. Densidade demográfica. Grau de urbanização. Índice de Gini. Taxa de analfabetismo. Taxa de desemprego. Taxa de pobreza. Taxa de policiais por 100 mil habitantes.	Presença de autocorrelação espacial. OLS: Significativos e positivos: densidade demográfica, grau de urbanização e taxa de pobreza. Significativos e negativos: policiais por 100 mil habitantes. O modelo explica 15% da variabilidade. SAR: Significativos e positivos: grau de urbanização e taxa de pobreza. O modelo explica 35% da variabilidade. SEM: Significativos e positivos: grau de urbanização e taxa de pobreza. Significativos e negativos: policiais por 100 mil habitantes. O modelo explica 11% da variabilidade. SDM: Significativos e positivos: grau de urbanização e taxa de pobreza. O modelo explica 40% da variabilidade.

Gomes et al (2017)	<p>Global Moran's</p> <p>Local Moran's</p> <p>OLS</p> <p>Modelo de Defasagem Espacial com erro de Média Móvel Espacial (SARMA)</p> <p>GWR</p>	Municípios do estado de Minas Gerais	<p>Taxa de crimes violentos: homicídio, homicídio tentado, estupro, roubo e roubo à mão armada.</p> <p>Taxa de crimes violentos contra o patrimônio: roubo e roubo à mão armada.</p> <p>Taxa de crimes violentos contra a pessoa: homicídio, homicídio tentado e estupro e taxa de homicídios.</p> <p>Por 100 mil habitantes.</p>	Existência de estabelecimento penitenciário ou prisional (ExUniPri), gasto per capita com infraestrutura (GasPCInfEst), gasto per capita com segurança pública (GasPCSegPub), gasto per capita com desenvolvimento econômico (GasPCDesEco), habitantes por policial civil ou militar (HabPol), população de 15 a 24 anos (Pop15a24), população de 25 a 29 anos (Pop25a29), renda per capita (RenPC), taxa de frequência ao ensino médio (TxFreEnsMed), densidade demográfica (DenPop) e taxa de armas apreendidas (TxArmApr).	<p>Presença de autocorrelação espacial nas variáveis dependentes.</p> <p>Modelo para crimes violentos:</p> <p>Significativos e positivos: ExUniPri, GasPcSegPub, Pop15a24, pop25a29 e RenPC.</p> <p>Significativos e negativos: GasPCInfEst, GasPCDesEco, HabPol e TxFreEnsMed.</p> <p>O modelo explicou 49% da variabilidade.</p> <p>Modelo para crimes contra o patrimônio:</p> <p>Significativos e positivos: ExUniPri, Pop15a24, Pop25a29 e RenPC.</p> <p>Significativos e negativos: GasPCInfEst, GasPCSegPub, GasPCDesEco, HabPol e TxFreEnsMed.</p> <p>O modelo explicou 56% da variabilidade.</p> <p>Modelo para crimes contra a pessoa:</p> <p>Significativos e positivos: ExUniPri e GasPCSegPub.</p> <p>Significativos e negativos: Pop25a29, RenPC e TxFreEnsMed.</p> <p>O modelo explicou 18% da variabilidade.</p> <p>Modelo para taxa de homicídio:</p> <p>Significativos e positivos: ExUniPri e Pop15a24.</p> <p>Significativos e negativos: Pop25a29, RenPC e TxFreEnsMed.</p> <p>O modelo explicou 16% da variabilidade.</p>
--------------------	---	--------------------------------------	---	---	---

Referência	Metodologia	Área de estudo	Variável Dependente (crime)	Variáveis Independentes	Resultados principais
Moreira & Fochezatto (2017)	Distribuição espacial do crime Global Moran's LISA Hot Spot OLS LM	Municípios do estado da Bahia	Índice de criminalidade gerado a partir de: homicídio doloso, lesão corporal seguida de morte, roubo com resultado de morte, tentativa de homicídio, estupro, roubo a ônibus, roubo de veículos e furto de veículos. Por 100 mil habitantes.	Densidade demográfica Vínculo per capita: total de indivíduos empregados no mercado formal de trabalho dividido pela população total de cada município. Matrícula per capita: total de alunos matriculados no Ensino Médio dividido pela população total de cada município. PIB per capita: O Produto Interno Bruto municipal dividido pela população total de cada município.	Presença de autocorrelação espacial. Relação positiva e significativa: densidade demográfica, matrícula per capita, PIB per capita. Relação não significativa: vínculo per capita. O modelo explicou 50% da variabilidade.
Saraiva et al (2017)	Distribuição espacial do crime Global Moran's LISA OLS LM	Municípios do estado do Rio Grande do Sul	Criminalidade agregada, considerando-se: homicídio doloso, furto, furto de veículo, roubo, roubo de veículo, latrocínio, extorsão, extorsão mediante sequestro, estelionato, delito relacionado a armas e munições, posse de entorpecentes e tráfico de entorpecentes. Por 100 mil habitantes.	Distribuição de renda: Theil-L, logaritmo neperiano da razão entre as médias aritméticas e geométricas da renda familiar per capita média. Urbanização: Percentagem da população da área urbana em relação à população total. Taxa de abandono do ensino fundamental. Taxa de abandono do ensino médio. Desordem dos núcleos familiares: Percentual de divórcios em relação à população total. Vínculo empregatício: Percentual de vínculos com até três meses de duração sobre o total dos vínculos.	Presença de autocorrelação espacial. Relação positiva e significativa: desordem dos núcleos familiares, vínculo empregatício, abandono do ensino médio, desigualdade de renda e urbanização. Relação não significativa: abandono do ensino fundamental. O modelo explicou 42% da variabilidade.

Referência	Metodologia	Área de estudo	Variável Dependente (crime)	Variáveis Independentes	Resultados principais
Bartz et al (2018)	Distribuição espacial do crime Global Moran's OLS	Conselhos Regionais de Desenvolvimento do estado do Rio Grande do Sul	Taxa de roubo por 100 mil habitantes. Taxa de furto por 100 mil habitantes. Taxa de homicídio por 100 mil habitantes.	Taxa de urbanização. Densidade populacional. Proporção de pessoas com superior completo. Proporção de pessoas que frequentaram creche. Proporção de residências com crianças onde nenhum dos moradores tinha ensino fundamental completo. PIB per capita. Renda per capita do 20% mais pobres. Renda per capita dos 10% mais ricos. Taxa de analfabetismo. Índice de GINI. Proporção de lares onde mulheres são chefes de família.	Autocorrelação espacial não significativa para todos os tipos de crime. Quatro modelos foram realizados. Modelo 1 (roubos): taxa de urbanização, densidade populacional, prop. sup. completo, prop. de pessoas que freq. creche, prop. Residências com crianças onde nenhum dos moradores tinha ensino fund. completo. O modelo 1 explicou 85% da variabilidade. Modelo 2 (homicídios): densidade populacional, renda 20% mais pobres, renda 10% mais ricos. O modelo 2 explicou 39% da variabilidade. Modelo 3 (homicídios): densidade populacional, PIB per capita, taxa analfabetismo. O modelo 3 explicou 53% da variabilidade. Modelo 4 (furtos): taxa de urbanização, prop. de lares onde mulheres são chefes de família, índice de GINI. O modelo 4 explicou 45% da variabilidade. Todas as variáveis independentes utilizadas nos modelos são significativas. As relações são todas positivas, fora prop. de pessoas que freq. creche e renda 20% mais pobres que apresentam direção negativa.

Tabela 2 - Variáveis independentes utilizadas em modelos de criminalidade no Brasil.

VARIÁVEL INDEPENDENTE	RESULTADOS
Taxa de analfabetismo	Lima et al (2005) relação positiva com a taxa de homicídios (percentual de pessoas com 15 anos ou mais que não são capazes de ler ou escrever). Batella (2008) correlação positiva para crimes contra o patrimônio e correlação negativa para crimes contra a pessoa. Sass et al (2016) sem relação significativa com homicídios. Bartz et al (2018) relação positiva com a taxa de homicídios.
Índice de pobreza	Lima et al (2005) relação positiva com a taxa de homicídios. Sass et al (2016) relação positiva com homicídios.
Desigualdade de renda	Batella (2008) sem correlação para crimes contra o patrimônio e correlação positiva baixa para crimes contra a pessoa. Oliveira (2008) relação positiva com furtos, roubos e homicídios. Saraiva et al (2017) relação positiva com a criminalidade agregada.
Taxa de frequência à escola	Batella (2008) correlação positiva para crimes contra o patrimônio e correlação negativa para crimes contra a pessoa. Gomes et al (2017) relação negativa com crimes violentos, crimes contra o patrimônio, crimes contra a pessoa e com a taxa de homicídio.
Percentual de pessoas de 15 ou mais analfabetas	Batella (2008) correlação positiva para crimes contra o patrimônio e correlação negativa para crimes contra a pessoa.
Desenvolvimento Humano (IDH)	Batella (2008) correlação positiva para crimes contra o patrimônio e correlação negativa para crimes contra a pessoa.
Riqueza	Batella (2008) correlação positiva para crimes contra o patrimônio e para crimes contra a pessoa.
Infra-estrutura	Batella (2008) correlação positiva para crimes contra o patrimônio e negativa para crimes contra a pessoa.
Educação	Batella (2008) correlação positiva para crimes contra o patrimônio e negativa para crimes contra a pessoa. Oliveira (2008) relação positiva com furtos, roubos e homicídios.

VARIÁVEL INDEPENDENTE	RESULTADOS
Estrutura populacional	Batella (2008) correlação positiva para crimes contra o patrimônio e para crimes contra a pessoa (População total, população de 15 a 24 anos de idade, densidade demográfica).
Imigração	Batella (2008) correlação positiva para crimes contra o patrimônio e para crimes contra a pessoa.
Renda dos mais pobres	Oliveira (2008) relação negativa com homicídios.
Renda dos mais ricos	Oliveira (2008) relação positiva com furtos e roubos.
Densidade demográfica	Oliveira (2008) relação positiva com furtos, roubos e homicídios. Sass et al (2016) relação positiva com homicídios. Moreira & Fochezatto (2017) relação positiva com o índice de criminalidade. Bartz et al (2018) relação positiva com roubos e homicídios.
Mulheres chefes de família	Oliveira (2008) relação positiva com furtos, roubos e homicídios. Bartz et al (2018) relação positiva com a taxa de furtos.
Habitações subnormais	Oliveira (2008) relação positiva roubos.
Grau de urbanização	Sass et al (2016) relação positiva com homicídios. Saraiva et al (2017) relação positiva com a criminalidade agregada. Bartz et al (2018) relação positiva com taxa de roubos e furtos.
Porcentagem de homens entre 15 e 24 anos na população	Sass et al (2016) sem relação significativa com homicídios.
Índice de Gini	Sass et al (2016) sem relação significativa com homicídios. Bartz et al (2018) relação positiva com a taxa de furtos.
Taxa de desemprego	Sass et al (2016) sem relação significativa com homicídios.
Taxa de policiais por 100 mil habitantes	Sass et al (2016) relação negativa com homicídios.
Existência de estabelecimento penitenciário	Gomes et al (2017) relação positiva com crimes violentos, crimes contra o patrimônio, crimes contra a pessoa e com a taxa de homicídio.
Gasto per capita com infraestrutura	Gomes et al (2017) relação negativa com crimes violentos e crimes contra o patrimônio.

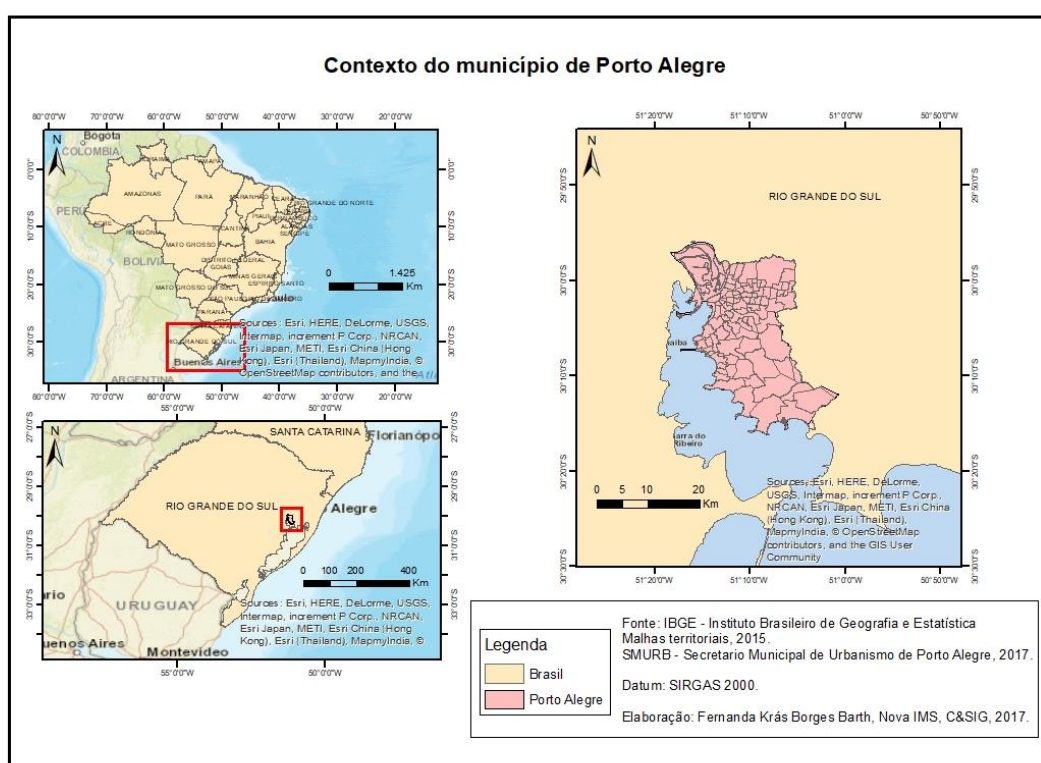
VARIÁVEL INDEPENDENTE	RESULTADOS
Gasto per capita com segurança pública	Gomes et al (2017) relação positiva com crimes violentos e crimes contra a pessoa, e relação negativa com crimes contra o patrimônio.
Gasto per capita com desenvolvimento econômico	Gomes et al (2017) relação negativa com crimes violentos e crimes contra o patrimônio, e relação positiva com crimes contra a pessoa.
Habitantes por policial civil ou militar	Gomes et al (2017) relação negativa com crimes violentos e crimes contra o patrimônio.
População de 15 a 24 anos	Gomes et al (2017) relação positiva com crimes violentos, crimes contra o patrimônio e com a taxa de homicídio.
População de 25 a 29	Gomes et al (2017) relação positiva com crimes violentos e crimes contra o patrimônio, e relação negativa com crimes contra a pessoa e a taxa de homicídio.
Renda per capita	Gomes et al (2017) relação positiva com crimes violentos e crimes contra o patrimônio, e relação negativa com crimes contra a pessoa e a taxa de homicídio.
Taxa de armas apreendidas	Gomes et al (2017) sem relação significativa com os crimes.
Vínculo empregatício per capita	Moreira & Fochezatto (2017) sem relação significativa.
Matrícula per capita	Moreira & Fochezatto (2017) relação positiva com o índice de criminalidade.
PIB per capita	Moreira & Fochezatto (2017) relação positiva com o índice de criminalidade. Bartz et al (2018) relação positiva com a taxa de homicídios.
Taxa de abandono do ensino fundamental	Saraiva et al (2017) sem relação significativa com a criminalidade agregada.
Percentual de divórcios em relação a pop. Total (desordem dos núcleos familiares)	Saraiva et al (2017) relação positiva com a criminalidade agregada.

VARIÁVEL INDEPENDENTE	RESULTADOS
Percentual de vínculos empregatícios com até três meses de duração sobre o total dos vínculos	Saraiva et al (2017) relação positiva com a criminalidade agregada.
Proporção de pessoas com superior completo	Bartz et al (2018) relação positiva com a taxa de roubos.
Proporção de pessoas que frequentam a creche	Bartz et al (2018) relação negativa com a taxa de roubos.
Proporção de residências com crianças onde nenhum dos moradores tinha ensino fundamental completo	Bartz et al (2018) relação positiva com a taxa de roubos.
Renda per capita dos 20% mais pobres	Bartz et al (2018) relação negativa com a taxa de homicídios.
Renda per capita dos 10% mais ricos	Bartz et al (2018) relação positiva com a taxa de homicídios.

3. REGIÃO DE ESTUDO

A área de estudo deste trabalho é o estado do Rio Grande do Sul situado ao sul do Brasil. O Rio Grande do Sul tem uma população estimada para 2019 de 11.377.239 distribuída entre 497 municípios (IBGE, 2019). Porto Alegre, capital do estado, é a 10ª cidade mais populosa do Brasil concentrando cerca de 1.500.000 habitantes, sendo a região metropolitana de Porto Alegre a quinta maior do país com 4.293.050 de habitantes, representando 2,1% da população brasileira. A localização do Rio Grande do Sul e da sua capital Porto Alegre seguem abaixo ilustradas (Figura 1).

Figura 1 - Mapa de contextualização do Rio Grande do Sul e sua capital Porto Alegre.

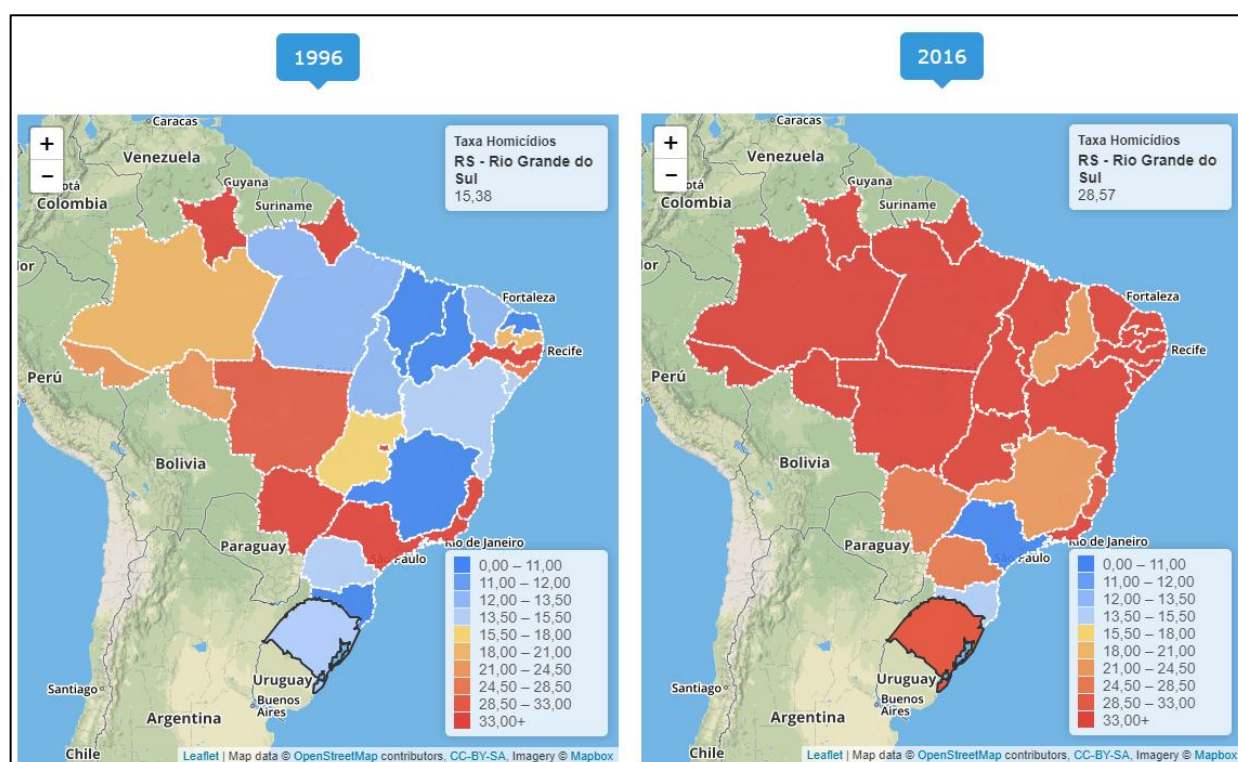


A criminalidade no estado do Rio Grande do Sul (RS), apesar de permanecer com valores baixos em relação aos restantes estados, vem aumentando, seguindo a tendência do país. Segundo Bartz et al (2018) o Rio Grande do Sul teve um crescimento maior do que o país em relação à taxa de homicídios (a cada 100 mil habitantes) da década 80 a 2010. No Brasil a taxa passou de 11,7 para 26,2, um aumento de 123% em 30 anos, e no Rio Grande do Sul, para o mesmo período, a taxa aumentou de 8,1

para 19,3, ou seja, um aumento de 138% em 30 anos. De acordo com Saraiva et al (2017), a profunda crise na segurança pública do estado se tornou uma grande preocupação entre os gaúchos (termo para denominar pessoas nascidas no estado), mas é um dos estados com menor efetivo policial per capita.

Segundo o IPEA (2019), e ilustrado na Figura 2, em 1996 a taxa de homicídios por 100 mil habitantes no Rio Grande do Sul era de 15,38, subindo gradativamente ano após ano, chegando a uma taxa de 28,57 em 2016, um aumento de 85,8% em 20 anos. Os roubos no Rio Grande do Sul, segundo dados da SSP-RS e da FEE, também tiveram um aumento: em 2008 a taxa de roubos por 100 mil habitantes foi de 570,5, em 2018 a taxa subiu para 632,9. A criminalidade, de forma geral, está aumentando no estado, porém, percebe-se que o aumento da criminalidade não é um fenômeno exclusivo do estado (Figura 2). A taxa de homicídio cresceu em quase todos os estados, com raras exceções como o estado de São Paulo.

Figura 2 - Taxa de Homicídios nos estados brasileiros.



Fonte: IPEA.

Ainda, de acordo com dados do IPEA de 2016, as cidades Viamão, Alvorada e Porto Alegre estão no ranking das 100 cidades mais violentas do Brasil, as três cidades fazem parte da região metropolitana de Porto Alegre. Segundo Bartz et al (2018), o crescimento da taxa de homicídios na região metropolitana de Porto Alegre entre 1980 e 2010 foi muito maior do que no país e no estado, registrando um aumento de 353,8%. Isto demonstra que o crime não possui uma distribuição homogênea no estado.

De acordo com Saraiva et al (2017) a criminalidade no Rio Grande do Sul ganha maior importância, não só pelo grave aumento, mas pela escassa literatura sobre o tema. Segundo os autores essa agenda de pesquisa é nova e pouco consolidada, sendo composta, sobretudo, por estudos realizados nos últimos 10 anos.

4. MATERIAIS E MÉTODOS

Para investigar os padrões da criminalidade no Rio Grande do Sul e suas causas foi utilizada uma abordagem de modelagem estatística e espacial. Os métodos aqui empregados, assim como as variáveis, têm sido utilizados nos estudos de criminalidade, conforme a revisão bibliográfica. Os modelos de regressão geograficamente ponderada (GWR – Geographically Weighted Regression) são modelos locais que permitem a modelação de fenômenos que exibem heterogeneidade espacial (Brunsdon, Fotheringham, & Charlton, 1996), como por exemplo a presença de agrupamentos ou hot spots. Segundo Fotheringham, Brunsdon, & Charlton (2002, p. 117), estes modelos também constituem uma solução alternativa para o problema dos termos de erro espacialmente autocorrelacionados.

Na literatura mais recente, foi proposta a modelação da criminalidade através de modelos GWR (e.g., Cahill & Mulligan, 2007; Stein, Conley, & Davis, 2016; Wang, Lee, & Williams, 2019), e através de modelos GWPR – Geographically Weighted Poisson Regression (e.g., Chen et al., 2017; Vilalta et al., 2019). Os modelos de regressão de Poisson permitem modelar dados de contagens e taxas, e assumem que a variável de estudo tem uma distribuição de Poisson. Por outro lado, quando a relação entre a variável dependente e as variáveis explicativas não é linear, os modelos de regressão linear múltipla estimados por Ordinary Least Squares (OLS; mínimos quadrados ordinários) e os modelos GWR não são adequados.

Após a recolha e pré-processamento dos dados, o desenvolvimento do trabalho de pesquisa foi dividido em quatro fases: (1) a análise exploratória; (2) a modelação de regressão global por OLS; (3) a modelação espacial local por GWR, e (4) a modelação com regressões de Poisson (global e local por GWPR).

A fase da análise exploratória dos dados foi feita com técnicas de estatística descritiva e de Exploratory Spatial Data Analysis (ESDA). Esta parte é crucial para entender como se comportam os dados, a sua distribuição de probabilidade e a sua distribuição espacial.

A fase da modelação foi realizada com a construção de modelos para a criminalidade, a análise de diagnósticos dos modelos testados e a comparação do

melhor modelo global obtido com o modelo local. A fase de modelação global por OLS é importante para a exploração das variáveis explicativas e sua significância, assim como para validar as teorias sobre os fatores causais da criminalidade. Para este efeito, foram investigados diversos modelos globais por OLS com diferentes variáveis potencialmente explicativas. Por fim, as variáveis incluídas no melhor modelo global obtido foram utilizadas no modelo local GWR, bem como nos modelos de regressão de Poisson (global e local por GWPR).

As secções seguintes descrevem mais detalhadamente os dados utilizados e cada uma das etapas da análise.

4.1 Recolha e o pré-processamento dos dados

Embora haja muitos tipos de crime, considerou-se que cada crime deve ser estudado individualmente, pois normalmente os fatores determinantes não são os mesmos, assim como a distribuição espacial não é igual. Portanto, não se levantou a hipótese de realizar uma taxa agregada dos crimes. Também, em função da limitação de tempo e facilidade de análise, não se poderia analisar todos os tipos de crime individualmente neste trabalho. Foram escolhidos, então, dois tipos de crime para se analisar individualmente: o homicídio e o roubo.

A proposta foi de retratar tanto um tipo de crime contra a pessoa quanto um tipo de crime contra o patrimônio. Apesar de muitos autores utilizarem apenas a taxa de homicídio como indicador da criminalidade, segundo Saraiva et al (2017) isso pode ser problemático, pois essa variável não engloba todas as formas de violência. O homicídio é um crime contra a pessoa, já o roubo é um crime contra o patrimônio. Teoricamente estes tipos de crimes se diferenciam em termos de motivação e causas, e a revisão de literatura reforça esta hipótese.

Portanto, as variáveis escolhidas para representar a criminalidade são a taxa de homicídio e a taxa de roubos por 100 mil habitantes. É habitual empregá-los em forma de taxa, pois, ao comparar diferentes unidades de território, as unidades com maior população tendem a apresentar maior ocorrência de crimes em números

absolutos, porém, isto não garante que sejam os locais mais violentos. Portanto, para poder compará-las, é necessário levar em conta a população.

As variáveis independentes foram selecionadas tendo em vista as variáveis já utilizadas na bibliografia e a abordagem de Oliveira (2008), que leva em conta a teoria econômica em conjunto da abordagem ecológica. Inicialmente haviam mais variáveis independentes, porém foram eliminadas a partir da matriz de correlação linear, afim de evitar multicolinearidade.

Todas as variáveis utilizadas no trabalho estão dispostas a seguir (Tabela 3). As variáveis independentes representam as seguintes dimensões (Tabela 4): educação, renda, ambiente (social e familiar), população, ocupação (trabalho) e qualidade de vida. O número absoluto de homicídios e roubos, assim como a estimativa da população total, foram utilizados para se obter as respectivas taxas.

As variáveis independentes foram adquiridas através dos dados do censo de 2010. Assume-se aqui, portanto, que o padrão das variáveis explicativas não sofreu grandes alterações entre 2010 e 2013.

Tabela 3 - Variáveis utilizadas: abreviatura, descrição, ano e fonte.

Abreviatura	Variável	Definição	Ano	Fonte
Homicídio	Número de homicídios	Número absoluto de homicídios para cada ano	2013, 2014, 2015, 2016 e 2017	Secretaria de Segurança Pública do estado do Rio Grande do Sul (SSP)
Roubos	Número de roubos	Número absoluto de roubos	2013, 2014, 2015, 2016 e 2017	Secretaria de Segurança Pública do estado do Rio Grande do Sul (SSP)
POP	População	Estimativa da população	2013, 2014, 2015, 2016 e 2017	Fundação de Economia e Estatística (FEE)
Taxa homic	Taxa de homicídio por 100 mil habitantes	Número absoluto de homicídios dividido pela população total e multiplicado por 100 mil	2013, 2014, 2015, 2016 e 2017	SSP e FEE

Taxa roubo	Taxa de roubo por 100 mil habitantes	Número absoluto de roubos dividido pela população total e multiplicado por 100 mil	2013, 2014, 2015, 2016 e 2017	SSP e FEE
T_ANALF15M	Taxa de analfabetismo - 15 anos ou mais	Razão entre a população de 15 anos ou mais de idade que não sabe ler nem escrever um bilhete simples e o total de pessoas nesta faixa etária multiplicado por 100	2010	Atlas do Desenvolvimento Humano do Brasil (IPEA)
T_FLMED	Taxa de frequência líquida ao médio	Razão entre o número de pessoas na faixa etária de 15 a 17 anos frequentando o ensino médio regular seriado e a população total dessa mesma faixa etária multiplicado por 100. As pessoas de 15 a 17 anos frequentando a 4ª série do ensino médio foram consideradas como já tendo concluído esse nível de ensino	2010	Atlas do Desenvolvimento Humano do Brasil (IPEA)
T_FREQ15A17	Taxa de atendimento escolar da população de 15 a 17 anos de idade	Razão entre a população de 15 a 17 anos de idade que estava frequentando a escola, em qualquer nível ou série e a população total nesta faixa etária multiplicado por 100	2010	Atlas do Desenvolvimento Humano do Brasil (IPEA)
T_FREQ18A24	Taxa de atendimento escolar da população de 18 a 24 anos de idade	Razão entre a população de 18 a 24 anos de idade que estava frequentando a escola, em qualquer nível ou série e a população total nesta faixa etária multiplicado por 100	2010	Atlas do Desenvolvimento Humano do Brasil (IPEA)
T_ATRASO_O_MED	% de 15 a 17 anos no médio sem atraso	Razão entre o número de pessoas de 15 a 17 anos frequentando o ensino médio regular seriado sem atraso idade-série e o número total de pessoas nessa faixa etária frequentando esse nível de ensino multiplicado por 100.	2010	Atlas do Desenvolvimento Humano do Brasil (IPEA)
GINI	Índice de Gini	Mede o grau de desigualdade existente na distribuição de indivíduos segundo a renda domiciliar per capita. Seu valor varia de 0, quando não há desigualdade (a renda domiciliar per capita de	2010	Atlas do Desenvolvimento Humano do Brasil (IPEA)

		todos os indivíduos tem o mesmo valor), a 1, quando a desigualdade é máxima (apenas um indivíduo detém toda a renda). O universo de indivíduos é limitado àqueles que vivem em domicílios particulares permanentes		
RENOCUP	Rendimento médio dos ocupados - 18 anos ou mais	Média dos rendimentos de todos os trabalhos das pessoas ocupadas de 18 anos ou mais de idade. Valores em reais de agosto de 2010	2010	Atlas do Desenvolvimento Humano do Brasil (IPEA)
PIND	Proporção de extremamente pobres	Proporção dos indivíduos com renda domiciliar <i>per capita</i> igual ou inferior a R\$ 70,00 mensais, em reais de agosto de 2010. O universo de indivíduos é limitado àqueles que vivem em domicílios particulares permanentes	2010	Atlas do Desenvolvimento Humano do Brasil (IPEA)
PREN20RICOS	Percentual da renda apropriada pelos 20% mais ricos	Percentual da renda total apropriada pelos indivíduos pertencentes ao quinto mais rico da distribuição dos indivíduos segundo a renda domiciliar per capita. O universo de indivíduos é limitado àqueles que vivem em domicílios particulares permanentes.	2010	Atlas do Desenvolvimento Humano do Brasil (IPEA)
T_DES18M	Taxa de desocupação - 18 anos ou mais	Percentual da população economicamente ativa (PEA) nessa faixa etária que estava desocupada, ou seja, que não estava ocupada na semana anterior à data do Censo, mas havia procurado trabalho ao longo do mês anterior à data dessa pesquisa	2010	Atlas do Desenvolvimento Humano do Brasil (IPEA)
T_NESTUDA_NTRAB_MMEIO	% de pessoas de 15 a 24 anos que não estudam nem trabalham e são vulneráveis à pobreza.	Razão entre as pessoas de 15 a 24 anos que não estudam nem trabalham e a população total nesta faixa etária multiplicado por 100. Define-se como vulneráveis à pobreza as pessoas que moram em domicílios com renda per	2010	Atlas do Desenvolvimento Humano do Brasil (IPEA)

		capita inferior a 1/2 salário mínimo de agosto de 2010. São considerados apenas os domicílios particulares permanentes		
T_DENS	Percentual da população que vive em domicílios com densidade superior a 2 pessoas por dormitório	Razão entre a população que vive em domicílios particulares permanentes com densidade superior a 2 e a população total residente em domicílios particulares permanentes multiplicado por 100. A densidade do domicílio é dada pela razão entre o total de moradores do domicílio e o número total de cômodos usados como dormitório	2010	Atlas do Desenvolvimento Humano do Brasil (IPEA)
T_MULCHEFEFIF014	% de mães chefes de família sem fundamental completo e com filhos menores de 15 anos	Razão entre o número de mulheres que são responsáveis pelo domicílio, não têm o ensino fundamental completo e têm pelo menos 1 filho de idade inferior a 15 anos morando no domicílio e o número total de mulheres chefes de família multiplicado por 100. São considerados apenas os domicílios particulares permanentes	2010	Atlas do Desenvolvimento Humano do Brasil (IPEA)
T_FUNDIN_TODOS	% de pessoas em domicílios em que ninguém tem fundamental completo	Razão entre as pessoas que vivem em domicílios em que nenhum dos moradores tem o ensino fundamental completo e a população total residente em domicílios particulares permanentes multiplicado por 100	2010	Atlas do Desenvolvimento Humano do Brasil (IPEA)
T_OCUPDESLOC_1	% de pessoas vulneráveis à pobreza e que gastam mais de uma hora até o trabalho	Razão entre as pessoas que vivem em domicílios vulneráveis à pobreza (com renda per capita inferior a 1/2 salário mínimo de agosto de 2010) e que gastam mais de uma hora em deslocamento até o local de trabalho e o total de pessoas ocupadas multiplicado por 100. São considerados apenas os domicílios particulares permanentes e as pessoas	2010	Atlas do Desenvolvimento Humano do Brasil (IPEA)

		com 18 anos de idade ou mais		
IDHM	Índice de Desenvolvimento Humano Municipal	Índice de Desenvolvimento Humano Municipal. Média geométrica dos índices das dimensões Renda, Educação e Longevidade, com pesos iguais	2010	Atlas do Desenvolvimento Humano do Brasil (IPEA)
T_M15A17CF	% de mulheres de 15 a 17 anos que tiveram filhos	Razão entre as mulheres de 15 a 17 anos de idade que tiveram filhos e o total de mulheres nesta faixa etária multiplicado por 100	2010	Atlas do Desenvolvimento Humano do Brasil (IPEA)
TAXAHOM15A24POP	Percentual de homens de 15 a 24 anos na população total	Número absoluto de homens de 15 a 24 anos dividido pela população total	2010	Atlas do Desenvolvimento Humano do Brasil (IPEA)

Tabela 4 - Legenda da Tabela 3: Grupos das variáveis independentes.

EDUCAÇÃO
RENDAS
AMBIENTE
POPULAÇÃO
OCUPAÇÃO
QUALIDADE VIDA

Após o pré-processamento dos dados no Excel, outro dado foi adquirido para dar continuidade ao trabalho: o shapefile do Rio Grande do Sul na projeção conforme cônica de Lambert, disponibilizado pela Universidade Federal do Rio Grande do Sul (UFRGS) na base cartográfica vetorial contínua do Rio Grande do Sul. Na sequência, retirou-se o município Pinto Bandeira da análise, pois este se emancipou em 2013, não havendo dados do censo de 2010 para o mesmo.

4.2 Análise exploratória

A análise exploratória foi dividida entre a estatística descritiva e o ESDA. Na fase da estatística descritiva optou-se por fazer uma análise dos dados de criminalidade de 2013 e dos anos seguintes até 2017, afim de observar a evolução da criminalidade desde o ano da modelação até os anos mais recentes.

Na fase do ESDA, foram aplicados os testes estatísticos de Moran Global e Moran Local e a análise de HotSpots. O Global Moran's I calcula o valor da estatística de Moran, o z-score e p-value para avaliar se o padrão expresso é agrupado, disperso ou aleatório. O índice de Moran varia entre -1 e 1, um valor positivo indica uma tendência de agrupamento (cluster), um valor negativo indica uma tendência de dispersão, e o valor 0 indica que o processo espacial subjacente é aleatório.

A estatística de Local Moran's I é utilizada para verificar se os padrões locais se diferem do padrão global, identificando se existem clusters e outliers espaciais estatisticamente significantes. A conceitualização do relacionamento espacial foi baseada em *contiguity edges only* com *standardization* por se tratarem de polígonos. Essa conceitualização leva em conta a ideia de que dois polígonos que compartilham fronteira têm uma interação espacial maior.

A análise de Hot Spots aplica a estatística de Getis-Ord G_i^* . Os resultados do z-score e do p-value identificam agrupamentos de valores muito elevados (hot spots) e agrupamentos de valores muito baixos (cold spots). Uma cidade com o valor muito elevado não será um hotspot se não tiver cercado de cidades com o valor, também, elevado. Esta análise é frequentemente utilizada para explorar padrões da criminalidade.

A análise espacial exploratória da taxa de homicídios por 100 mil habitantes nos municípios do RS em 2013 (secção 5.2) revelou que esta variável não apresenta autocorrelação espacial significativa (estatística de Moran Global igual a 0,0284 com p-value=0,2626), e que apenas um dos municípios corresponde a um hot spot com um grau de confiança de 95% (estatística Getis-Ord G_i^*). Uma vez que não há evidência nem de autocorrelação espacial nem de heterogeneidade significativas no padrão espacial da taxa de homicídios, não se justifica a sua análise através de modelos de regressão espacial.

4.3 Modelos OLS

Análise de regressão permite que se modele, examine e explore relações espaciais para melhor entender os fatores que estão por trás dos padrões observados, e assim prever padrões futuros. A regressão Ordinary Least Squares (OLS, em

português: mínimos quadrados ordinários) estima os parâmetros de uma regressão linear com o princípio de mínimos quadrados: minimizando a soma dos quadrados das diferenças entre os valores observados da variável dependente e os valores previstos pela função linear. O modelo OLS é considerado global, pois cada variável tem apenas um parâmetro associado. As relações são consideradas espacialmente estacionárias, ou seja, se assume que a relação entre as variáveis explicativas e as observações de crimes são as mesmas em todos os locais (Cahill & Mulligan, 2007).

O método de estimação OLS para modelos lineares é um dos mais comuns. Isto se dá porque o modelo é bastante confiável se ele atende os pressupostos para regressões lineares. Os principais pressupostos são:

- **Os regressores são fixos:** As variáveis independentes não são estocásticas, ou seja, são resultados de um processo não aleatório.
- **Sem multicolinearidade:** As variáveis independentes não possuem correlação linear entre si.
- **Exogeneidade estrita:** O erro é aleatório com média 0.
- **Homoscedasticidade:** A variância do erro é constante.
- **Sem autocorrelação:** Os erros são independentes.
- **Distribuição normal:** Os erros têm uma distribuição normal.

Caso alguma dessas premissas não seja verdadeira, o método pode gerar resultados sub-ótimos ou enviesados. Para avaliar o modelo e testar as premissas são utilizadas algumas técnicas de diagnóstico listadas a seguir (nos testes estatísticos considerou-se um nível de significância de 5%, exceto indicação explícita em contrário):

- **Jarque-Bera:** testa a normalidade dos resíduos (Bai & Ng, 2005). Os resíduos são previsões acima ou abaixo do esperado, eles devem refletir ruído aleatório. É esperado que os resíduos sejam distribuídos em uma normal e que tenham um padrão espacial aleatório. Se o teste Jarque-Bera rejeitar a hipótese nula

da normalidade ($p\text{-value} \leq 0,05$), os resíduos não estão distribuídos em uma normal, portanto o modelo poderá estar enviesado.

- **Koenker test:** testa a heteroscedasticidade dos resíduos (Koenker, Bassett, & Jr, 1982). Quando o teste Koenker rejeita essa hipótese nula ($p\text{-value} \leq 0,05$) as relações modeladas não são consistentes, isso pode ser causado por heteroscedasticidade ou pela não estacionaridade das relações. Nestes casos, apenas se pode confiar no Robust t-test para determinar a significância dos coeficientes (fornecido pelo Software ArcGIS).
- **Variance Inflation Factor (VIF):** serve para avaliar a presença de multicolinearidade. Valores abaixo do valor normalmente utilizado de 10 geralmente indicam não-multicolinearidade entre as variáveis explicativas do modelo (O'Brien, 2007). Se o valor for acima, significa que existe redundância entre algumas variáveis explicativas, isso resulta em um modelo com viés.
- **Robust t-test:** testa a significância das variáveis explanatórias ($p\text{-value} \leq 0,05$), serve como alternativa ao usual t-test que não pode ser confiável quando há heteroscedasticidade.
- **Wald test:** avalia a significância do modelo (Wald, 1943), serve, também, como alternativa ao usual F-test que não pode ser confiável quando há heteroscedasticidade. Ou seja, quando o teste Koenker for significativo, é necessário utilizar o teste Wald para avaliar a significância geral do modelo. Se a hipótese nula do teste de Wald não for rejeitada ($p\text{-value} > 0,05$) significa que a equação de regressão não explica a variação na variável dependente.

As variáveis foram selecionadas com base na não-multicolinearidade, previamente testada e reafirmada com o teste VIF, e com base no teste-t robusto (Robust Probability): apenas variáveis que se apresentaram significativas neste teste permaneceram no modelo. Para testar a autocorrelação espacial dos resíduos, aplicou-se o teste estatístico de Moran Global aos mesmos.

4.4 Modelo GWR

O modelo GWR é um método de regressão local e espacial que permite que as relações modeladas possam variar pela área de estudo. Segundo Fotheringham et al (Fotheringham et al., 2002), desenvolvedores do modelo GWR, o problema de utilizar parâmetros globais é que estes são essencialmente médias espaciais que podem esconder uma grande quantidade de informações sobre o processo em estudo.

Este modelo é uma extensão do modelo de regressão linear global. Enquanto o modelo de regressão linear global estima um parâmetro para cada variável independente, no modelo GWR um parâmetro é estimado para cada variável independente em cada local na área de estudo. Assume-se que os coeficientes são funções determinísticas de outra variável, neste caso da localização no espaço (Fotheringham et al., 2002).

Um método para comparar a performance dos modelos OLS e GWR é a comparação dos coeficientes de determinação ajustado (AdjR2) e do critério de informação de Akaike corrigido (AICc; Hurvich, Simonoff, & Tsai, 1998). O valor do AdjR2 varia entre 0 e 1 e expressa o quanto da variação da variável dependente foi explicado pelo modelo, sendo 1 o modelo explica 100% da variabilidade da variável dependente. O AICc é um valor que pode ser utilizado para medir a performance de um modelo quando comparado com outro; um modelo com menor AICc é considerado melhor. Em resumo, um AdjR2 maior significa um maior potencial explicativo, enquanto um AICc menor indica aprimoramento no modelo.

A implementação do GWR neste trabalho foi calibrada a partir do kernel adaptativo e do método AICc para encontrar o número ótimo de vizinhos, de forma automática no software ArcGIS. O kernel utilizado é uma função quase-Gaussiana que pondera a observação j na regressão localizada em i através do peso w_{ij} definido por

$$w_{ij} = \exp \left[- (d_{ij}/b)^2 \right]$$

onde d_{ij} é a distância entre o ponto da regressão i e o ponto vizinho com o valor observado j , e b é o parâmetro *bandwidth* que varia em função do número ótimo de vizinhos.

As variáveis independentes são as mesmas selecionadas previamente a partir do modelo global OLS. As técnicas de diagnóstico utilizadas contemplam o teste de Moran Global aplicado aos resíduos para conferir a autocorrelação espacial dos resíduos, e o Condition Number (também designado na literatura por Condition Index) que deverá ser inferior a 30; caso contrário, há evidência de multicolinearidade entre as variáveis independentes. Caso o modelo GWR se revele apropriado, será também elaborado um mapa dos erros padrão para avaliar a precisão dos coeficientes locais estimados.

Para maiores detalhes sobre o modelo GWR veja-se Fotheringham et al (Fotheringham et al., 2002).

4.5 Modelos de Poisson

Embora o uso de dados de criminalidade em modelos lineares seja frequente na literatura, de acordo com Chen et al (2017), quando o número de crimes é usado como variável dependente, o pressuposto da normalidade é violado, pois o número de crimes é uma variável de contagem. Consequentemente as estimativas serão enviesadas e há subestimação dos erros padrão dos coeficientes de regressão. Osgood (2000) já questionava o uso de OLS para taxas de crime, quando é comum que muitas das unidades (e.g., cidades) apresentem um número baixo de eventos. Por um lado, taxas baseadas em populações de menor dimensão tenderão a apresentar maiores erros padrão. Por outro lado, não se pode assumir que a distribuição dos resíduos seja normal quando as contagens de crime são pequenas. Como o menor número de crimes que se pode obter é zero, então a distribuição dos resíduos deverá tornar-se cada vez mais assimétrica quando as taxas se aproximam deste limite.

Portanto, as regressões de Poisson são utilizadas para lidar com variáveis não-negativas como dados de contagem e taxas. É assumido que a variável dependente tem uma distribuição de Poisson, e que o logaritmo do seu valor esperado pode ser modelado por uma combinação linear de parâmetros desconhecidos. A Poisson foi originalmente proposta como distribuição de probabilidade para análise de taxas de condenação na França, por Siméon Denis Poisson no seu trabalho "*Probabilité des*

jugements en matière criminelle et en matière civile, précédées des règles générales du calcul des probabilités" (Paris, France: Bachelier, 1837).

A *Geographically weighted Poisson regression* (GWPR) é uma extensão da GWR proposta por (Nakaya, Fotheringham, Brunson, & Charlton, 2005) para lidar com este tipo de distribuição.

A regressão de Poisson assume que a variável dependente tem uma distribuição de Poisson, e assume que o logaritmo do seu valor esperado pode ser modelado por uma combinação linear de parâmetros desconhecidos. Isto implica que se assume que o valor médio (valor esperado da variável) é igual à variância. Na prática, observa-se frequentemente que a variância é muito superior à média. Neste caso, diz-se que estamos perante um problema de sobre-dispersão (*overdispersion*) em relação à distribuição de Poisson (i.e., a variância observada é superior à variância esperada).

O modelo de regressão de Poisson faz parte da família de modelos lineares generalizados, e não contém um termo de erro aleatório na sua forma pura. Sendo assim, a função de máxima verosimilhança não se baseia na densidade dos erros aleatórios e sim na distribuição da própria variável dependente. Por este motivo, ao contrário da regressão linear, não é necessário estimar a variância dos erros a partir dos resíduos, já que não existe termo de erro no modelo (Anselin, 2004).

Tal como nos modelos de regressão linear, a forma de comparar os modelos de Poisson é através do AICc mais baixo. Todavia, não é possível comparar o AICc de um modelo linear com o AICc de um modelo de Poisson, pois o AICc baseia-se na função da máxima verosimilhança e esta, como visto anteriormente, é diferente para estes modelos. Para avaliar a performance geral do modelo utiliza-se a percentagem de *deviance* explicada.

Serão utilizados os seguintes modelos de Poisson:

- **Modelo de regressão de Poisson global:** consiste numa única equação para toda a região de estudo, assim como o OLS;
- **Modelo GWPR (regressão de Poisson local):** GWR adaptado a regressão de Poisson, ou seja, considera as relações não-estacionárias;
- **Modelo GWPR semi-paramétrico (S-GWPR):** permite que algumas variáveis sejam consideradas globais e outras locais (como no GWR).

A implementação destes modelos foi efetuada através do programa GWR4, a partir da opção Local-To-Global onde são automaticamente exploradas as variáveis para os modelos S-GWPR (para mais detalhes, veja-se Nakaya, Fotheringham, Charlton, & Brunsdon, 2009). O kernel utilizado nos modelos foi o Gaussiano adaptativo para evitar ou reduzir problemas de sobre-dispersão.

5. RESULTADOS

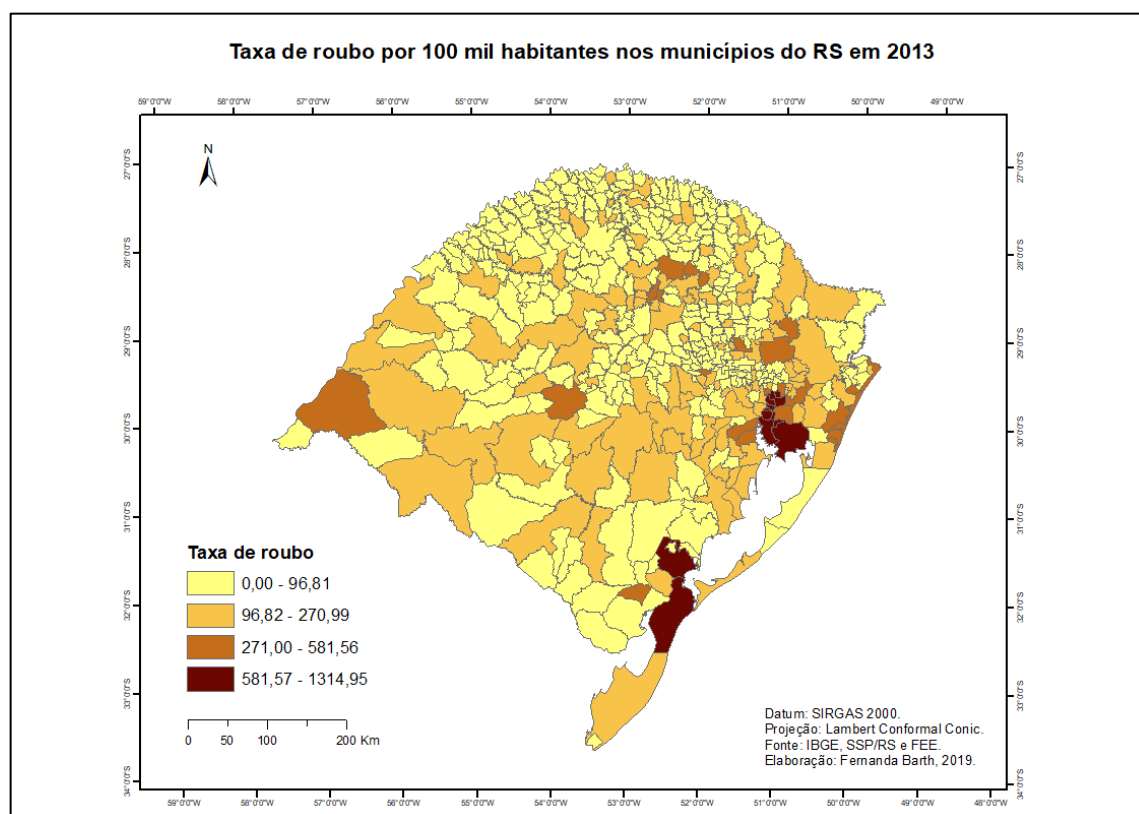
Os resultados foram apresentados por tipo de análise e por tipo de crime: roubos e homicídios. Primeiramente são expostos os resultados da análise exploratória dos dados com estatística descritiva, seguido da análise exploratória de dados espaciais (ESDA). Após, são apresentados os resultados da modelação da criminalidade associada aos roubos.

5.1 Análise exploratória da criminalidade associada a roubos

Antes de seguir para a análise de regressão, é importante entender a distribuição dos próprios dados e, também, sua distribuição espacial. Nesta análise exploratória breve estes padrões são investigados através de estatísticas descritivas e do ESDA.

O mapa a seguir (Figura 3) apresenta a distribuição da taxa de roubo pelo estado no ano de 2013 (as classes da legenda foram definidas usando *natural breaks*). Ao observá-lo é possível identificar algumas particularidades no padrão da taxa de roubo. Por exemplo, muitos municípios com uma baixa taxa de roubos por todo estado, destacando-se a região norte. Também fica em evidência a concentração de taxas altas de roubo em poucas cidades, nomeadamente na região metropolitana de Porto Alegre: Alvorada, Cachoeirinha, Campo Bom, Canos, Esteio, Novo Hamburgo, Porto Alegre, São Leopoldo, Sapucaia do Sul e Viamão; e os municípios ao Sul: Pelotas e Rio Grande.

Figura 3 - Mapa da distribuição da taxa de roubo no estado em 2013.



5.1.1 Estatísticas descritivas

Em uma primeira análise exploratória dos dados foram dispostas as medidas descritivas do número de roubos para os anos de 2013 a 2017 (Tabela 5). O mínimo se refere ao valor mínimo registrado, o Q1 representa o primeiro quartil, o Q2 representa o segundo quartil ou mediana, o Q3 representa o terceiro quartil e o máximo se refere ao valor máximo registrado.

Ao observar as medianas e a os máximos constata-se que a maior parte das cidades do Rio Grande do Sul têm uma ocorrência baixa de roubos e algumas poucas cidades concentram elevados números de ocorrências. Esta informação está de acordo com o mapa anterior (Figura 3). Também se constata que os números absolutos de roubos têm aumentado a cada ano. Em 4 anos houve um aumento em torno de 66% na média dos roubos.

Tabela 5 - Medidas descritivas do número anual de roubos de 2013 a 2017 no Rio Grande do Sul.

Ano	Mínimo	Q1	Mediana (Q2)	Q3	Máximo	Média
2013	0	1	3	15	19301	106,27
2014	0	1	3	14	24454	125,11
2015	0	1	4	19	31343	161,94
2016	0	1	5	21	34825	176,02
2017	0	1	4	24	34900	176,86

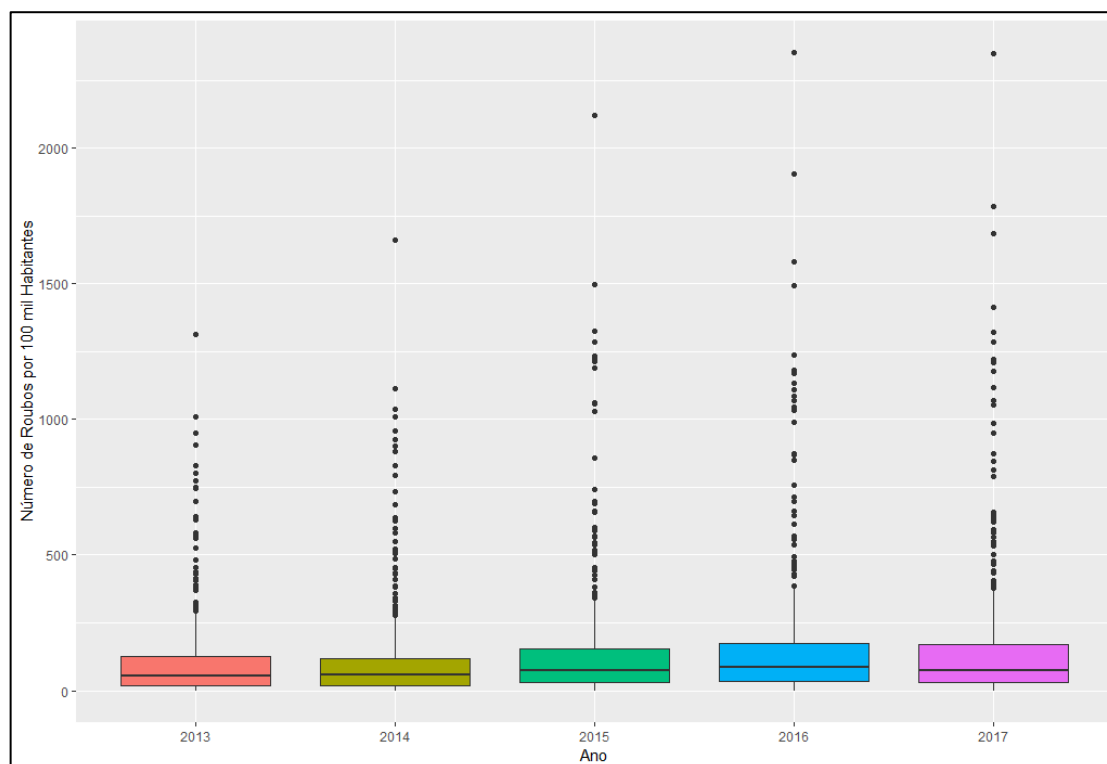
Já ao observar as medidas descritivas da taxa de roubo por 100 mil habitantes (Tabela 6), percebe-se uma redução na diferença entre os valores medianos e os valores máximos. Isso se dá porque, ao trabalhar com a taxa, retira-se o efeito da concentração populacional de grandes centros urbanos e as diferenças entre as cidades diminuem. Entretanto a diferença entre as medianas e os valores máximos para cada ano ainda é grande, ou seja, há poucas cidades que concentram altas taxas de roubo e o resto das cidades registram poucos roubos por habitantes. A média da taxa de roubo aumentou em torno de 52% em 4 anos.

Tabela 6 - Medidas descritivas da taxa anual de roubos por 100 mil habitantes de 2013 a 2017 no Rio Grande do Sul.

Ano	Mínimo	Q1	Mediana (Q2)	Q3	Máximo	Média
2013	0	16,60	53,63	126,64	1314,95	101,92
2014	0	19,74	57,01	119,56	1660,73	108,05
2015	0	31,18	73,86	154,87	2122,26	139,43
2016	0	34,49	87,12	173,24	2351,42	157,69
2017	0	31,85	75,31	168,30	2350,26	154,42

Uma técnica interessante para a análise exploratória descritiva é o uso de Box plots, estes permitem uma boa visualização da distribuição dos dados. A tabela a seguir (Tabela 7) demonstra a existência de valores extremos e o aumento desses anualmente, enquanto a maior parte da distribuição (até 75% dos dados) permanece próxima ao 0. Isto mostra a grande concentração de altos valores em poucas cidades.

Tabela 7 - Box plot do número de roubos por 100 mil habitantes de 2013 a 2017 no Rio Grande do Sul.



5.1.2 Autocorrelação espacial

O resultado do teste estatístico de Global Moran's I (Figura 4) confirma que a taxa de roubo tem autocorrelação espacial (positiva), com chance menor que 1% dos padrões espaciais serem aleatórios. Com um índice de Moran de 0,50, as taxas de roubo tendem a agrupar-se no estado, ou seja, cidades com altas taxas são próximas de cidades com altas taxas e vice-versa.

Já o teste de Local Moran's I (Figura 5) revela a presença de clusters de altos valores na região metropolitana de Porto Alegre, ao norte (Mato Castelhano e Passo Fundo), em alguns municípios do Litoral Norte do Estado (Osório, Tramandaí, Xangri-lá, Balneário Pinhal, Capão da Canoa e Cidreira) e ao sul (Pelotas, Rio Grande e Capão do Leão). Junto ao cluster *High-High* (i.e., valores elevados correlacionam-se significativamente com valores elevados vizinhos) ao Sul encontra-se o município Arroio do Padre que se apresenta como um Outlier espacial *Low-High* (i.e., valores baixos correlacionam-se significativamente com valores elevados vizinhos) com baixa taxa de roubos.

Figura 4 - Relatório do teste de Global Moran's I para taxa de roubos por 100 mil habitantes em 2013 no estado.

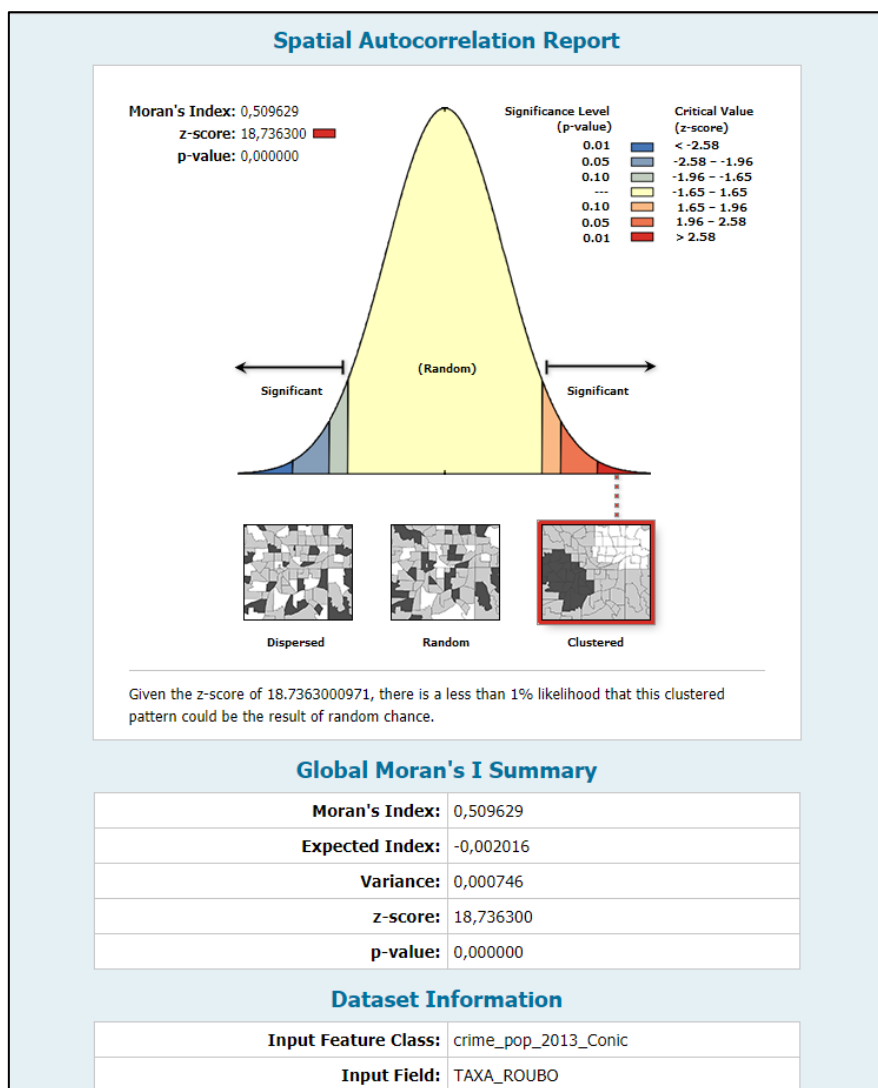
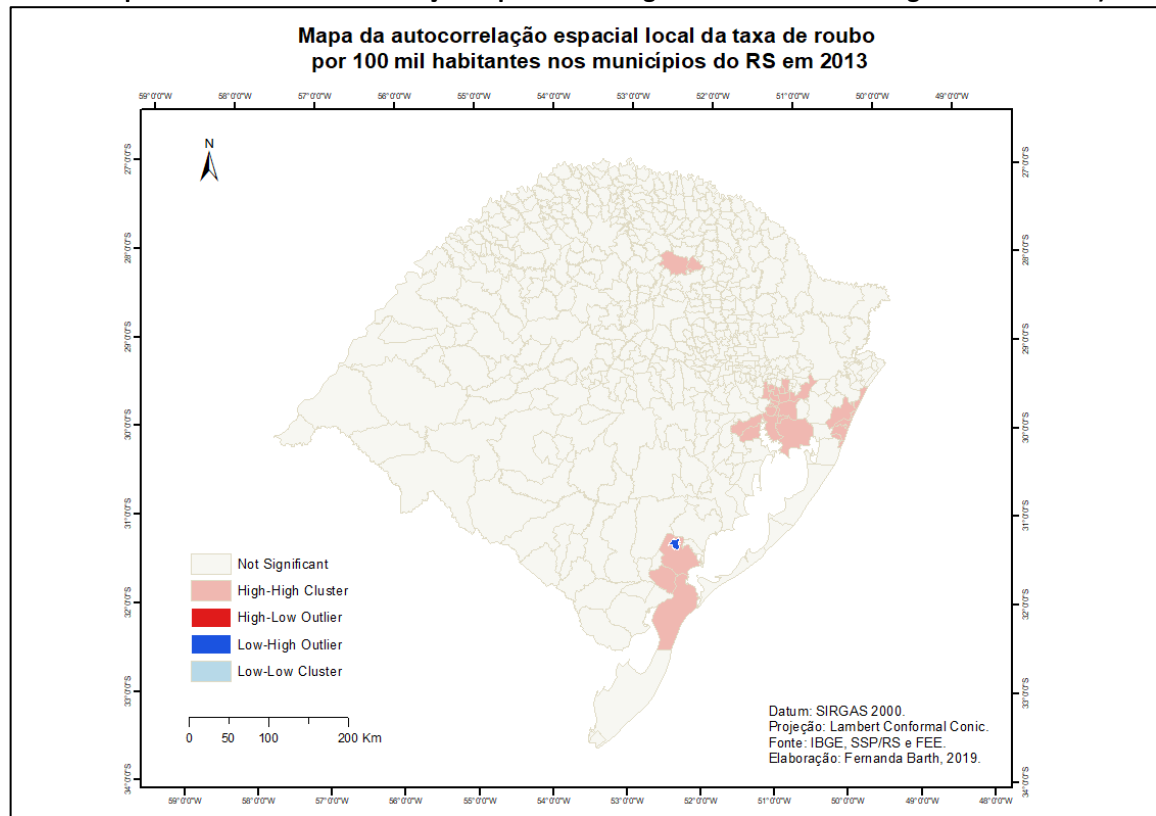


Figura 5 - Mapa do teste de Local Moran's I da taxa de roubo no estado em 2013 (os clusters e outliers espaciais indicam autocorrelação espacial local significativa ao nível de significância de 5%).

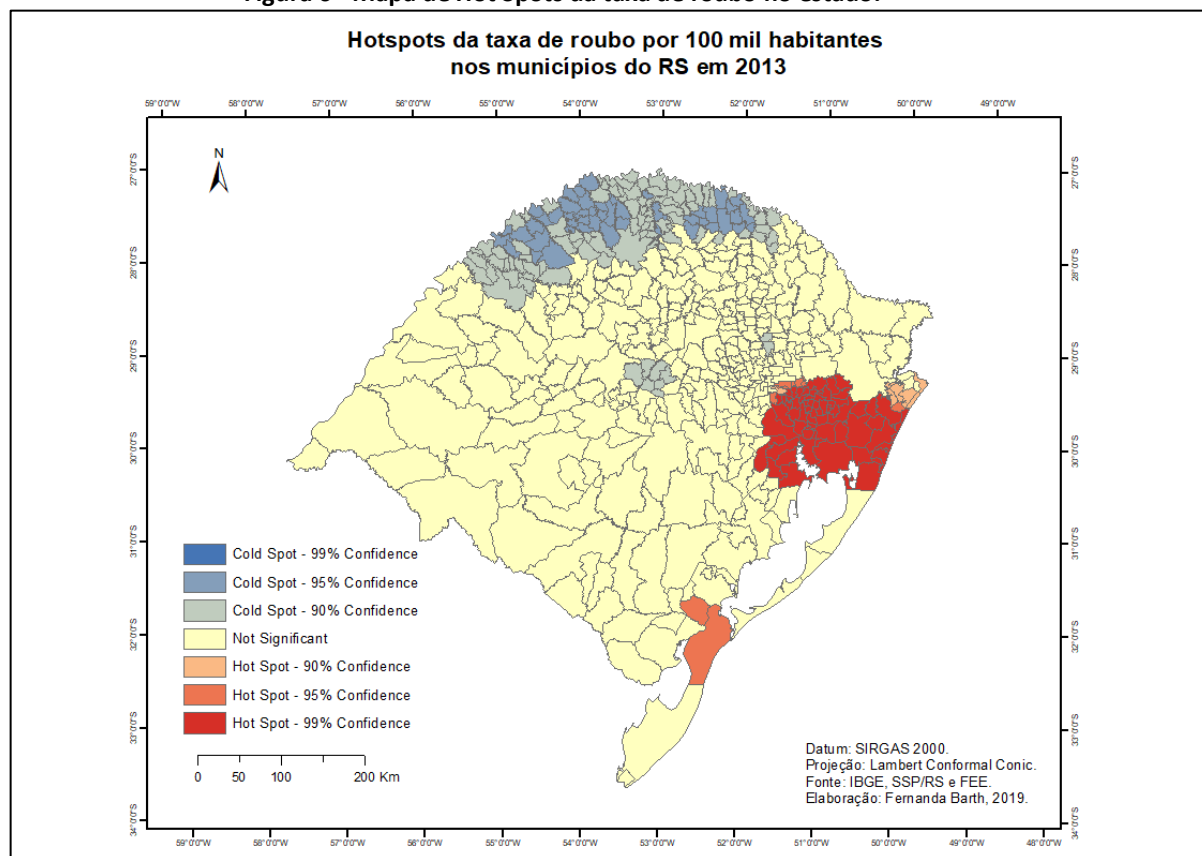


Este resultado condiz com a análise descritiva prévia que revela que poucas cidades concentram valores mais elevados. Estes locais de clusters positivos significativos são candidatos para políticas mais direcionadas de redução da criminalidade associada a roubos.

5.1.3 Hot Spots

A análise de Hot Spots revela agrupamentos de municípios com altas taxas de roubo na área metropolitana de Porto Alegre em conjunto com a região do litoral norte e mais ao Sul, em Rio Grande e Capão do Leão (Figura 6). O norte do estado apresenta um agrupamento de mais de 100 municípios com baixas taxas de roubos. Novamente destacam-se o litoral norte, a região metropolitana de Porto Alegre e ao sul o município de Rio Grande como regiões de maior criminalidade associada a roubos.

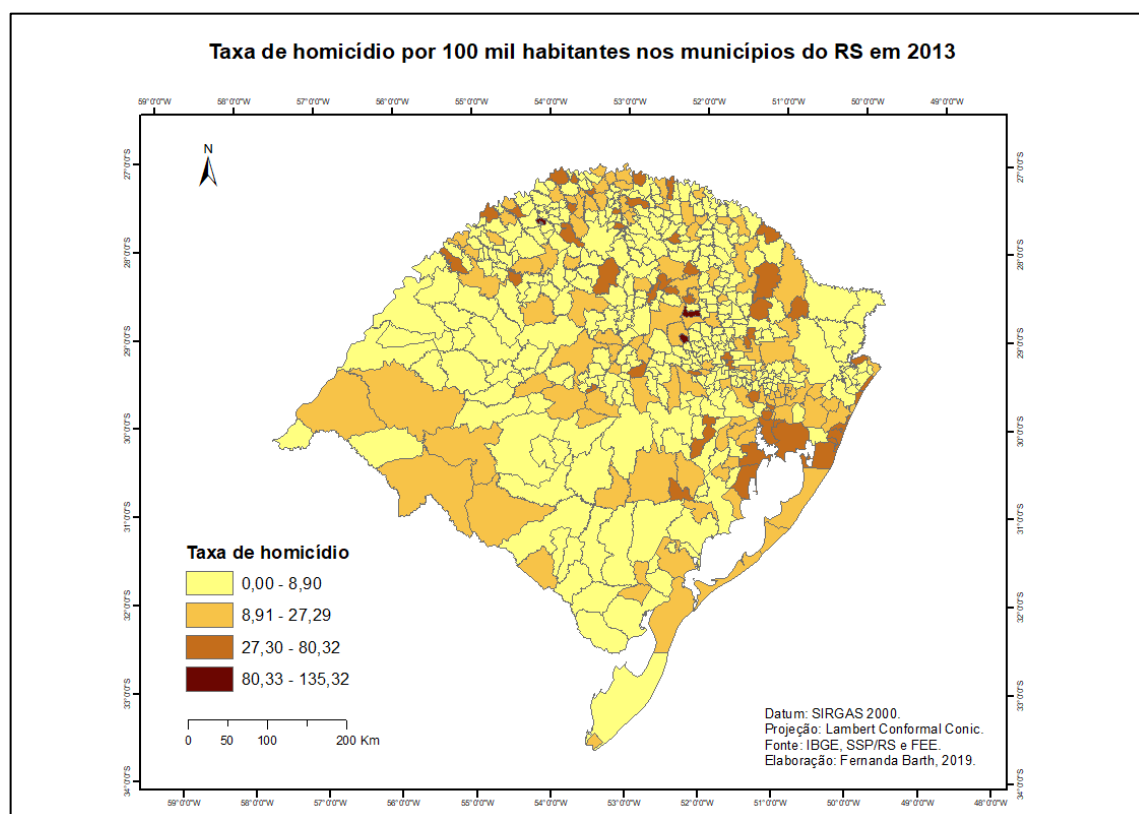
Figura 6 - Mapa de Hot Spots da taxa de roubo no estado.



5.2 Análise exploratória da criminalidade associada a homicídios

Nesta seção foi realizada a análise exploratória associada a homicídios. Na sequência, o mapa da distribuição da taxa de homicídios no estado para o ano de 2013 (Figura 7) não parece revelar nenhum padrão espacial específico. As taxas de homicídio parecem se distribuir de forma aleatória no estado. Nas próximas seções foram aplicados testes estatísticos para investigar de fato se há algum padrão ou se o padrão é aleatório.

Figura 7 - Mapa da distribuição da taxa de homicídio no estado em 2013.



5.2.1 Estatísticas descritivas

As medidas descritivas do número absoluto de homicídios (

Tabela 8) expõem grande disparidade entre os terceiros quartis e os valores máximos, o que indica valores extremos (outliers). A mediana é zero tanto para homicídios absolutos quanto para a taxa de homicídios (Tabela 9) ou seja, mais da metade das cidades não possui nenhum registro de homicídio para estes anos. O homicídio é muito menos frequente do que o roubo e o ele é mais concentrado em algumas cidades (outliers). Ao menos metade das cidades não registraram nenhum homicídio nos anos observados e pelo menos 75% dos municípios registraram, no máximo, apenas 2 homicídios por ano.

Tabela 8 - Medidas descritivas do número anual de homicídios no Rio Grande do Sul.

Ano	Mínimo	Q1	Mediana (Q2)	Q3	Máximo	Média
2013	0	0	0	1	463	3,88
2014	0	0	0	2	574	4,73
2015	0	0	0	2	587	4,89
2016	0	0	0	2	705	5,25
2017	0	0	0	2	581	5,37

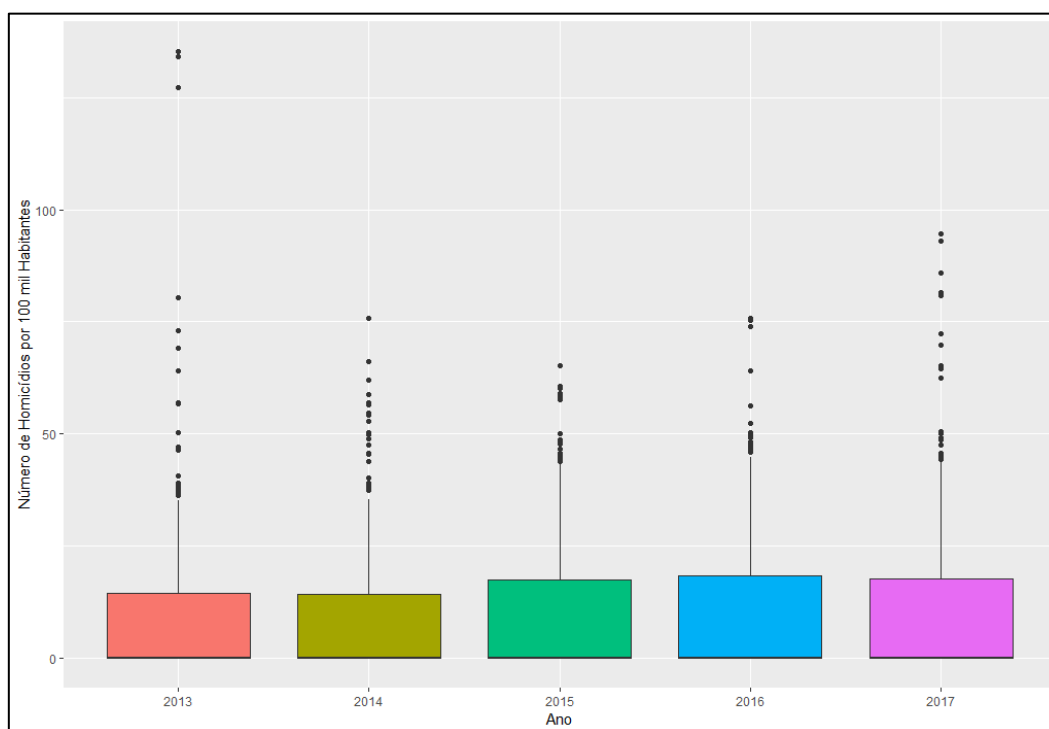
Novamente, como nos roubos, a divisão pela população (Tabela 9) ameniza, mas não elimina a presença de Outliers. Constata-se que a média dos homicídios vem aumentando, porém é bastante baixa.

Tabela 9 - Medidas descritivas da taxa de homicídios por 100 mil habitantes no Rio Grande do Sul.

Ano	Mínimo	Q1	Mediana (Q2)	Q3	Máximo	Média
2013	0	0	0	14,46	135,32	9,10
2014	0	0	0	14,21	75,84	9,12
2015	0	0	0	17,39	65,17	10,04
2016	0	0	0	18,21	75,80	10,62
2017	0	0	0	17,61	94,70	10,61

O Boxplot a seguir (Figura 8) evidencia as informações já analisadas nas tabelas anteriores. Ao observá-lo nota-se que o primeiro quartil é igual a mediana (zero), ou seja, há muitos municípios onde não foram registrados homicídios. Também constata-se que em 2013 houveram três municípios que se destacam muito nos valores extremos, não havendo o mesmo padrão para os anos seguintes. Isto pode ser devido a eventos atípicos ou pode até mesmo ser um erro na base de dados.

Figura 8 - Boxplot da taxa anual de homicídios por 100 mil habitantes.



5.2.2 Autocorrelação espacial

Segundo os resultados do teste de Moran Global (Figura 9) a taxa de homicídio não apresentou autocorrelação espacial significativa na área de estudo para um nível de significância de 5% ($p\text{-value} > 0,05$), assim como em Oliveira (2008) e Bartz et al (2018). Portanto, com índice de Moran próximo a 0, a distribuição espacial da taxa de homicídio é aleatória no estado para o ano de 2013.

Para investigar a existência de padrões locais de autocorrelação foi aplicado o teste de Moran Local. O resultado (Figura 10) revela que não há autocorrelação local significativa na área de estudo. Há apenas ocorrência de cluster de valores altos entre Porto Alegre e Alvorada, e mais ao norte, a ocorrência de Outliers espaciais positivos (valores altos cercados por valores baixos) em São José do Inhacora e São José do Herval, ambos registraram 3 homicídios no ano de 2013.

Portanto, os padrões de autocorrelação globais e locais da taxa de homicídio no Rio Grande do Sul para o ano de 2013 são aleatórios ou pouco relevantes.

Figura 9 - Relatório do teste Global Moran's I para taxa de homicídios por 100 mil habitantes.

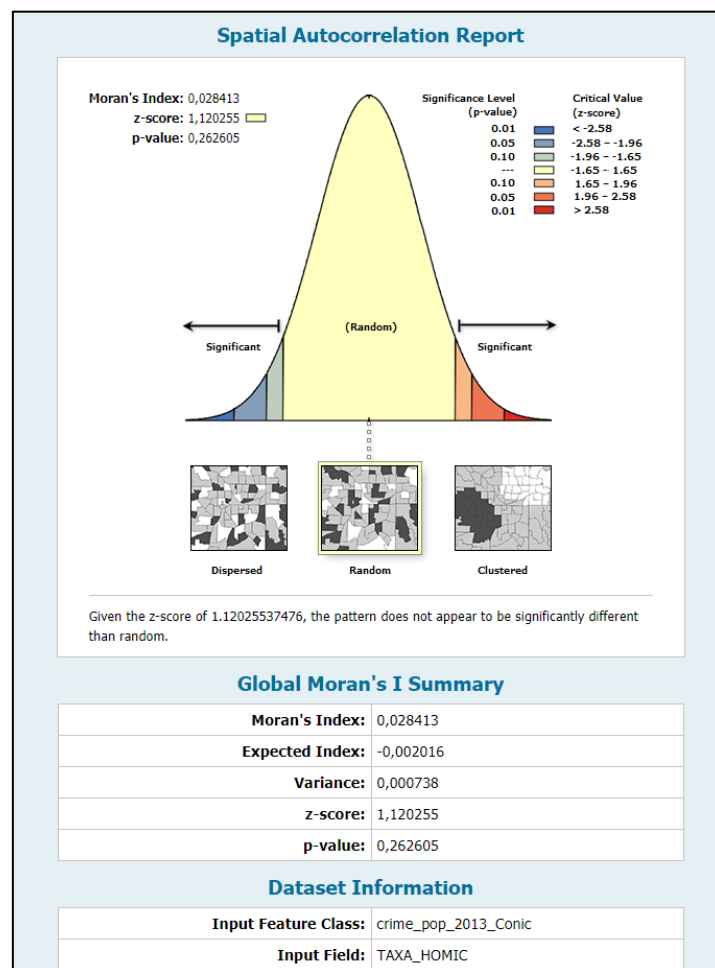
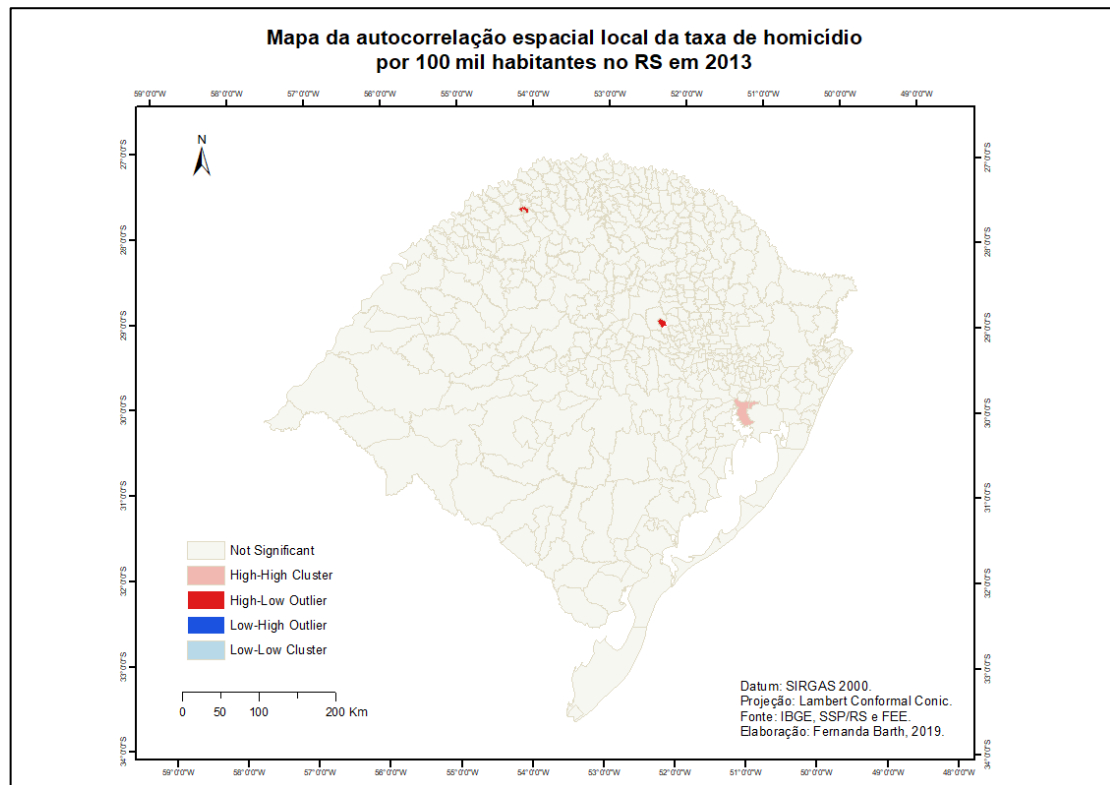


Figura 10 - Mapa do teste de Local Moran's I da taxa de homicídio no estado em 2013 (os clusters e outliers espaciais indicam autocorrelação espacial significativa ao nível de significância de 5%).

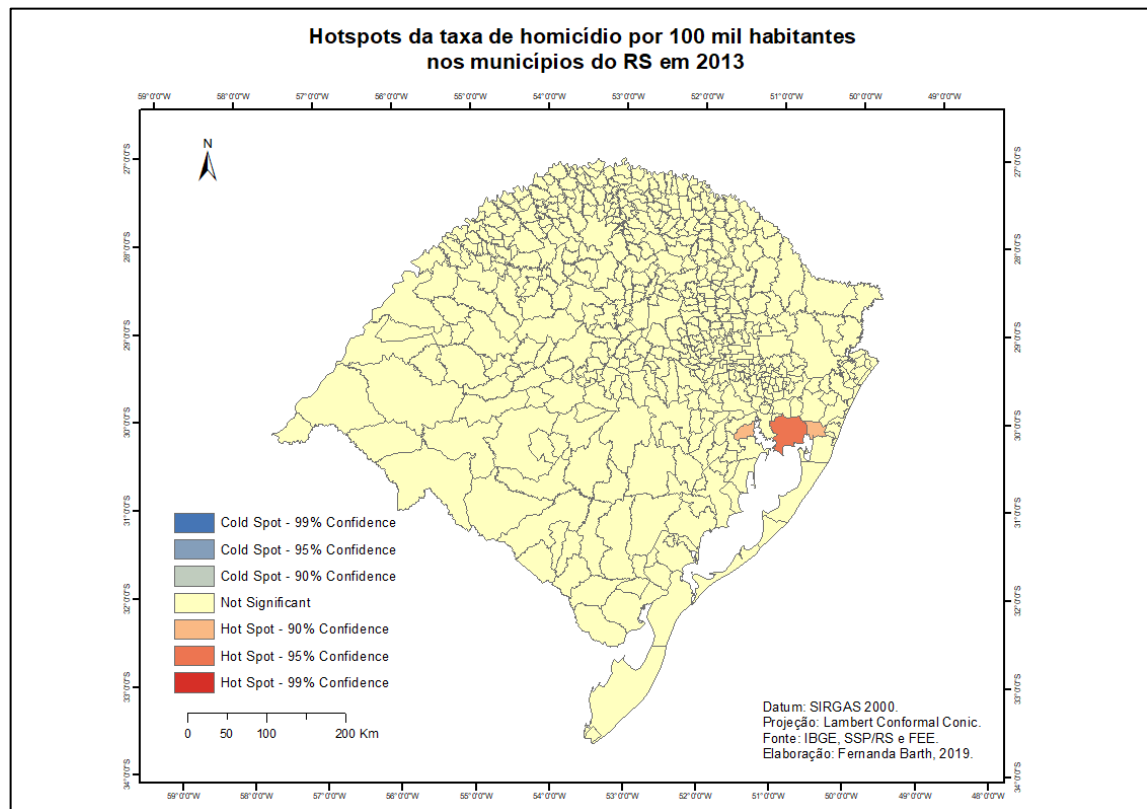


5.2.3 Hot Spots

A análise de Hot Spots (Figura 11) revela apenas um agrupamento significativo para valores elevados da taxa de homicídios no estado para o ano de 2013. Este resultado reforça a evidência de que o padrão espacial das taxas de homicídio não exhibe heterogeneidade significativa na maior parte da região de estudo.

Não havendo evidência de autocorrelação espacial, nem de não-estacionaridade relevante no padrão espacial, não se justifica a análise da taxa de homicídio através de modelos de regressão espacial, como já mencionado no capítulo de Materiais e Métodos.

Figura 11 - Mapa de Hot Spots da taxa de homicídios no estado em 2013.



5.3 Relação entre a criminalidade e potenciais variáveis explicativas

Um método para investigar possíveis relações entre variáveis é observar a correlação linear entre elas através de matrizes de correlação e gráficos de dispersão. A correlação linear varia de -1 a 1, sendo 1 a correlação positiva perfeita e -1 correlação negativa perfeita. Ou seja, se a correlação linear entre duas variáveis é positiva, quando uma aumenta a outra também aumenta; a correlação negativa significa que quando uma variável aumenta a outra diminui. O gráfico de dispersão é uma forma mais visual para observar o tipo de relação e a magnitude da relação entre variáveis: quanto mais próximos os pontos no gráfico em torno de uma reta, mais correlacionadas linearmente são as variáveis.

Ao avaliar a correlação linear entre os crimes (roubo e homicídio) e a população, o gráfico de dispersão (Figura 12) indica que há correlação linear positiva alta entre o número de homicídios, de roubos e a população, pois quando um aumenta os outros aumentam também. Já o gráfico de dispersão seguinte (Figura 13) leva em conta as

taxas, de homicídio e de roubo, por 100 mil habitantes ao invés do número absoluto. O resultado demonstra pouca correlação entre as taxas de roubo e de homicídio, a população tem correlação linear com a taxa de roubos, mas não aparenta ter correlação linear com a taxa de homicídios.

Figura 12 - Gráfico de dispersão do número absoluto de homicídios, número de roubos e população em 1 milhão para os anos de 2013 a 2017.

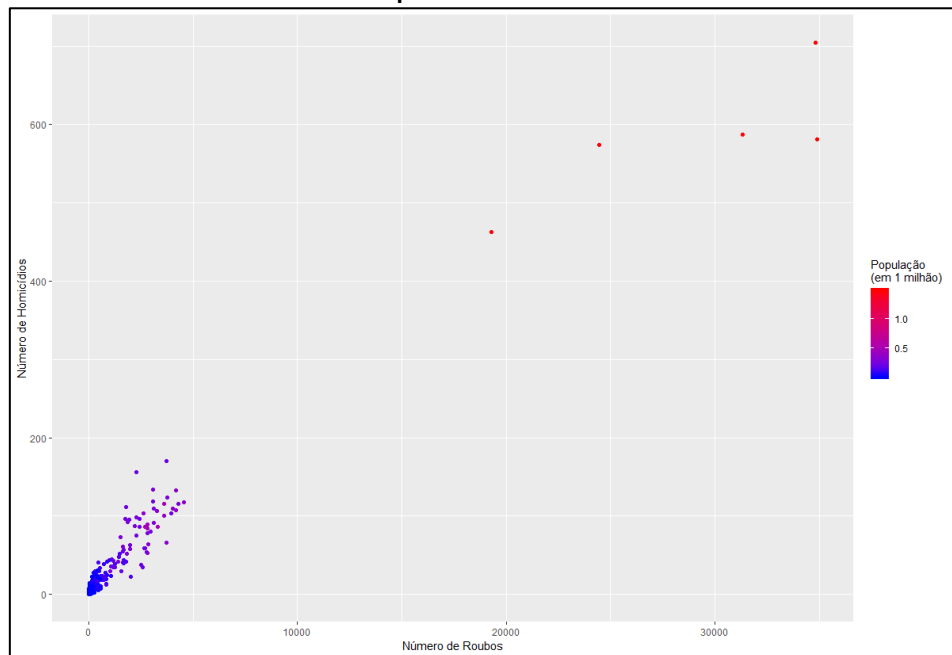
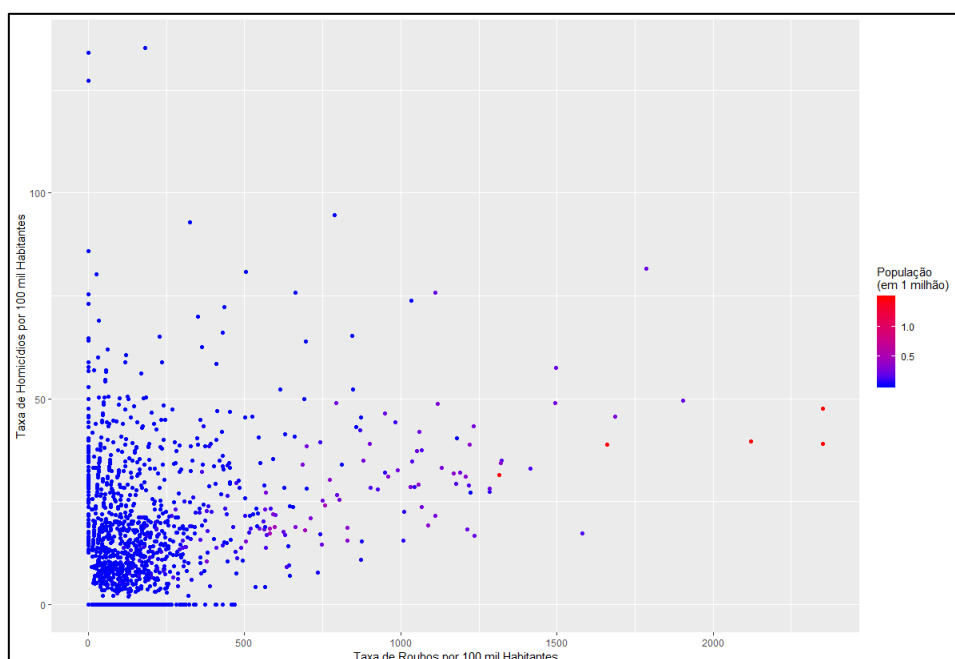


Figura 13 - Gráfico de dispersão da taxa de homicídio e de roubo por 100 mil habitantes e da população em 1 milhão, para os anos de 2013 a 2017.



Ao analisar a matriz de correlação linear entre homicídio, roubo, taxa de roubo, taxa de homicídio e a população (Tabela 10) confirma-se a altíssima correlação linear positiva entre a população e roubo, a população e homicídio e entre homicídio e roubo. Este resultado era esperado, pois quanto maior a população mais ocorrências de crimes. Já as taxas têm uma correlação linear bem menor entre si e entre a população, embora a correlação linear entre a taxa de roubo e a população é alta.

Tabela 10 - Matriz de correlação linear entre o homicídio, o roubo, a taxa de homicídio, a taxa de roubo e a população para os anos de 2013 a 2017 no estado.

	Taxa Homic.	Taxa Roubo	Pop.	Homic.	Roubo
Taxa Homic.	1				
Taxa Roubo	0,4	1			
Pop.	0,21	0,69	1		
Homic.	0,25	0,67	0,96	1	
Roubo	0,16	0,59	0,92	0,97	1

A seguir (Tabela 11) está disposta a matriz de correlação linear entre todas as variáveis independentes (2010) utilizadas, as taxas de roubo e de homicídio (2013), com intuito de ter uma visão ampla sobre todas as relações.

Tabela 11 - Matriz de correlação linear entre todas variáveis independentes e as variáveis dependentes (em negrito)

	<i>PREN20RICOS</i>	<i>GINI</i>	<i>RENOCUP</i>	<i>PIND</i>	<i>T_DES18M</i>	<i>T_NESTUD</i>	<i>T_DENS</i>	<i>T_MULCHEFEFIF014</i>	<i>T_FUNDIN_TODOS</i>	<i>T_OCUPDESLOC_1</i>	<i>IDHM</i>	<i>T_M15A17CF</i>	<i>TAXAHOMEM15A24POP</i>	<i>POP Estimada</i>	<i>Taxa_homic</i>	<i>Taxa_roubo</i>
<i>PREN20RICOS</i>	1,00															
<i>GINI</i>	0,98	1,00														
<i>RENOCUP</i>	0,30	0,19	1,00													
<i>PIND</i>	0,34	0,45	-0,53	1,00												
<i>T_DES18M</i>	0,37	0,35	0,20	0,04	1,00											
<i>T_NESTUDA_NTRAB_MMEIO</i>	0,43	0,49	-0,31	0,62	0,46	1,00										
<i>T_DENS</i>	0,38	0,43	-0,17	0,46	0,54	0,65	1,00									
<i>T_MULCHEFEFIF014</i>	0,16	0,18	-0,13	0,18	0,46	0,43	0,53	1,00								
<i>T_FUNDIN_TODOS</i>	-0,02	0,07	-0,68	0,48	-0,28	0,33	0,15	0,12	1,00							
<i>T_OCUPDESLOC_1</i>	0,24	0,27	-0,13	0,25	0,30	0,40	0,47	0,31	0,15	1,00						
<i>IDHM</i>	0,00	-0,11	0,77	-0,65	0,01	-0,57	-0,46	-0,34	-0,85	-0,35	1,00					
<i>T_M15A17CF</i>	0,19	0,22	-0,03	0,21	0,24	0,35	0,46	0,30	0,10	0,19	-0,24	1,00				
<i>TAXAHOMEM15A24POP</i>	0,05	0,03	0,20	-0,02	0,15	0,00	0,27	0,22	-0,28	0,15	0,07	0,14	1,00			
<i>POP Estimada</i>	0,16	0,14	0,39	-0,11	0,28	-0,02	0,17	0,09	-0,30	0,10	0,21	0,05	0,12	1,00		
<i>Taxa_homic</i>	0,09	0,10	0,04	0,05	0,14	0,07	0,16	0,09	-0,03	0,10	-0,02	0,13	0,00	0,15	1,00	
<i>Taxa_roubo</i>	0,13	0,10	0,41	-0,19	0,46	0,01	0,29	0,28	-0,40	0,17	0,24	0,11	0,20	0,68	0,24	1,00

Os resultados da correlação linear entre a taxa de homicídio e as variáveis independentes reforçam a decisão de não utilizar nos modelos de regressão, pois a correlação é muito baixa, ou quase inexistente, entre elas. Para a taxa de roubo, a correlação linear mais alta é entre a variável população, depois, respectivamente, entre a variável "Taxa de desocupação - 18 anos ou mais" (T_DES18M; 0,46), "Rendimento médio dos ocupados - 18 anos ou mais" (RENOCUP; 0,41) e "% de pessoas em domicílios em que ninguém tem fundamental completo" (T_FUNDIN_TODOS; -0,40).

Há algumas variáveis independentes que têm uma correlação forte entre si. O IDHM (Índice de desenvolvimento humano municipal), por exemplo, tem correlação alta com muitas variáveis. Isto acontece provavelmente porque o IDH é uma média geométrica das dimensões renda, educação e saúde, portanto resume muitas variáveis em um só índice. Este comportamento pode não ser muito bom para um modelo de regressão, pois ao utilizar esta variável teríamos de deixar de utilizar outras (para evitar multicolinearidade) que podem representar melhor as especificidades locais. De fato, a variável IDHM tem uma correlação linear mais baixa com a variável dependente (taxa de roubo) do que outras variáveis que tem correlação alta com ela. Outra correlação altíssima, quase perfeita, é encontrada entre as variáveis GINI (Índice de Gini) e PREN20RICO (Percentual da renda apropriada pelos 20% mais ricos). Isto poderá ser pelo fato de ambas variáveis serem indicadores de desigualdade.

A escolha de manter algumas variáveis correlacionadas permitiu investigar modelos diferentes e averiguar quais delas têm maior poder explicativo para o roubo. Em um modelo as variáveis independentes se unem para explicar a variável dependente, portanto a correlação linear é apenas um passo para ajudar a selecionar essas variáveis.

A seguir são expostas as correlações lineares da taxa de roubo (variável dependente) com todas variáveis independentes. Os gráficos de dispersão estão apresentados em anexo.

- **População estimada:** A variável população tem a correlação linear mais forte (0,677) com a variável dependente. Essa forte relação é visível no gráfico de dispersão. Embora alguns autores prefiram trabalhar com a densidade

demográfica, optou-se por se trabalhar diretamente com a população. Ambas descrevem uma história parecida quanto à criminalidade: quanto mais pessoas em um determinado local, mais crimes. Segundo Oliveira (2008) e Gomes et al (2017), os roubos ocorrem com maior frequência em áreas urbanas, onde há concentração de população, pois aumenta os benefícios e diminui a probabilidade da punição, assim como os custos de execução e planejamento. Isso seria por causa do anonimato do criminoso e da maior quantidade de vítimas potenciais em regiões densas.

- **Percentual de homens de 15 a 24 anos na população total (TAXAHOMEM15A24POP):** Sass et al (2016) comentam que as causas e os efeitos da criminalidade são mais intensos em faixas etárias mais jovens, principalmente masculina. Segundo os mesmos, essas faixas etárias são mais afetadas pela renda e desemprego. Porém, nos seus resultados a variável não se mostrou significativa no modelo para homicídios. Gomes et al (2017) indicam que a juventude reflete conflitos e desigualdades sociais do país, sendo essa uma fase de oportunidades. Quanto à taxa de roubo, a TAXAHOMEM15A24POP tem uma correlação linear relativamente baixa (0,205) e positiva, o gráfico da dispersão demonstra que aparentemente não existe uma relação linear entre essas variáveis.
- **Percentual de mulheres de 15 a 17 anos que tiveram filhos (T_M15A17CF):** Essa variável foi escolhida para averiguar a relação da gravidez na adolescência com a taxa de roubos. Oliveira (2008) recomenda políticas de redução da gravidez na adolescência como medidas para reduzir a desordem familiar (neste trabalho representada pela variável T_MULCHEFEFIF014). A correlação linear com a taxa de roubos, no entanto, é baixa (0,105) e o gráfico de dispersão exhibe pontos bastante dispersos, portanto parece não haver relação linear entre essas variáveis.
- **Índice de Desenvolvimento Humano Municipal (IDHM):** O IDH é uma variável que costuma ser empregada para representar a qualidade de vida em geral

(renda, educação e saúde). Batella (2008) e Lima et al (2005) utilizaram-na encontrando, respectivamente, correlação linear positiva para crimes contra o patrimônio e negativa para crimes contra pessoa, e correlação linear positiva com a taxa de homicídio. Lima et al (2005) acabaram por não a usar nos modelos de regressão, pois causaria multicolinearidade. O IDHM demonstrou uma baixa correlação linear (0,242) com a taxa de roubos.

- **Percentual de pessoas vulneráveis à pobreza e que gastam mais de uma hora até o trabalho (OCUPDESLOC):** Essa variável, assim como o IDHM, foi considerada como representante da qualidade de vida, mas também pode ser considerada como uma representação do ambiente (vulnerável à pobreza) e à oportunidade de emprego, visto que neste caso o indivíduo vulnerável à pobreza se sujeita a horas de deslocamento para poder trabalhar. Os resultados Figura permitem concluir que a correlação linear desta variável com a taxa de roubo é muito baixa (0,170).
- **Percentual de pessoas em domicílios em que ninguém tem fundamental completo (T_FUNDIN_TODOS):** Conforme Saraiva et al (2017), aspectos do ambiente histórico social do indivíduo, como ambiente familiar, exposição de violência e pobreza na infância, oportunidades de trabalho no mercado legal e um amplo conjunto de variáveis socioeconômicas e demográficas, são fundamentais na determinação do comportamento criminoso. Oliveira (2008) explica que os problemas no microssistema² dos indivíduos afetarão seu desenvolvimento, inclusive o desenvolvimento dos valores morais (custo moral). Esses microssistemas juntos são considerados no presente trabalho como ambiente. Tanto a variável T_FUNDIN_TODOS como as próximas se encaixam nessa abordagem. A correlação linear entre T_FUNDIN_TODOS e a taxa de roubo é relativamente baixa e negativa (-0,396). No gráfico de

² Oliveira (2008) explica que a abordagem ecológica insere os indivíduos em quatro sistemas concêntricos com suas interconexões. O microssistema são o conjunto de atividades, papéis e relações interpessoais experienciados pelo indivíduo em um ambiente específico. O conjunto de microssistemas forma o mesossistema que seria o conjunto de relações construídas ao longo da vida.

dispersão é perceptível a relação negativa (mas não linear), esta direção não era a prevista. Há a hipótese de que a relação desta variável com a taxa de roubo poderá variar no espaço, sendo de esperar que seja significativamente positiva em determinadas zonas e poderá não apresentar uma relação significativa em outras zonas.

- **Percentual de mães chefes de família sem fundamental completo e com filhos menores de 15 anos (T_MULCHEFIF014):** Essa variável é muito utilizada na bibliografia tendo apresentado relação direta com a criminalidade, conceitualmente é chamada de "desorganização familiar". Segundo Oliveira (2008), seria uma representação do papel da família no desenvolvimento do indivíduo, e o resultado positivo em relação ao crime se explicaria porque famílias monoparentais, em média, são mais vulneráveis em função da redução de renda familiar e o sobrecarregamento da mãe (por ausência paterna) que pode acarretar em problemas de criação dos filhos. O resultado indica que a correlação linear entre essa variável e a taxa de roubo é relativamente baixa (0,284).
- **Percentual da população que vive em domicílios com densidade superior a 2 pessoas por dormitório (T_DENS):** Essa variável representa o ambiente de moradia, com intuito de investigar se existe relação entre domicílios superlotados e a taxa de roubo. A superlotação domiciliar pode indicar problemas de renda e pode gerar problemas de privacidade e conforto (Givisiez & Oliveira, 2013). No gráfico de dispersão é perceptível uma tendência parabólica dos pontos, que explica uma correlação linear relativamente baixa (0,290).
- **Percentual de pessoas de 15 a 24 anos que não estudam nem trabalham e são vulneráveis à pobreza (T_NESTUDA_NTRAB_MMEIO):** O desemprego e a escolaridade são variáveis clássicas comumente correlacionadas à criminalidade (W. Batella, 2008). Esta variável representa indivíduos jovens que não possuem nem vínculo escolar nem empregatício e são vulneráveis à

pobreza. Teoricamente estes jovens com pouca oportunidade de ingressar em trabalhos lícitos e, necessitando de renda, poderiam adentrar no mercado ilícito. Os resultados revelam que a variável T_NESTUDA_NTRAB_MMEIO não tem correlação linear com a taxa de roubo (coeficiente de correlação linear de 0,009).

- **Taxa de desocupação - 18 anos ou mais (T_DES18M):** Indivíduos de 18 anos ou mais que estão à procura de trabalho, mas não conseguem, podem ser potenciais candidatos para o mercado ilícito e atividades criminosas. Segundo Batella (2008), o desemprego pode ter consequências como estresse, quebra da estrutura familiar e hábitos danosos à saúde, sendo estes fatores agravantes das taxas de criminalidade. O gráfico de dispersão não demonstra uma nítida distribuição dos pontos, porém a correlação linear é moderada e positiva (0,459), ou seja, quanto maior a taxa de desocupação de indivíduos com 18 anos ou mais, maior a taxa de roubo.
- **Proporção de indivíduos extremamente pobres (PIND):** A correlação linear entre PIND e taxa de roubo é igual a -0,192, sendo, portanto, baixa. Este sinal não era esperado, embora uma explicação possível é de que não é propriamente a proporção de indivíduos extremamente pobres na população que tem relação com a criminalidade e sim a desigualdade.
- **Rendimento médio dos ocupados - 18 anos ou mais (RENOCUP):** Segundo Oliveira (2008), é esperado que a renda tenha um sinal ambíguo, pois tem efeitos tanto nos benefícios do crime quanto nos custos de oportunidade. Conforme os resultados, a correlação linear entre RENOCUP é moderada e positiva (0,414). Esse resultado pode indicar que regiões com maior média de rendimento têm mais benefícios. Ou também, um rendimento médio alto não necessariamente é bem distribuído, pode ser uma região onde há muita concentração de renda, neste caso é também importante averiguar índices de desigualdades.

- **Índice de GINI:** O índice de GINI é uma medida de desigualdade de renda, onde 0 representa a equidade de renda. Sass et al (2016) e Bartz et al (2018) usaram o índice de GINI para representar a desigualdade de renda nos seus modelos de criminalidade. Para Sass et al (2016) a variável não apresentou significância no modelo de homicídios, porém para Bartz et al (2018) apresentou relação positiva significativa com a taxa de furtos. No presente trabalho, a correlação entre o índice de GINI e a taxa de roubo se revelou bastante baixa (0,098).
- **Percentual da Renda apropriada pelos 20% mais ricos (PREN20RICOS):** Essa é outra variável alternativa capaz de representar a concentração de renda e que se espera que tenha uma relação positiva com a taxa de roubo. Os resultados apresentam correlação linear baixa e positiva (0,125) entre PREN20RICOS e a taxa de roubo. Embora baixa, PREN20RICOS apresenta um resultado melhor do que o GINI. No entanto, a relação aparenta ser não linear.
- **Percentual de pessoas de 15 a 17 anos no ensino médio sem atraso (T_ATRASO_0_MED):** A educação tem um papel extremamente importante na construção da moral do indivíduo, das suas relações sociais, e ainda aumenta as chances de ingresso no mercado formal, assim como possibilita salários maiores (Oliveira, 2008; Saraiva et al., 2017). Ao observar o gráfico de dispersão não parece haver relação linear entre as variáveis. A correlação linear é baixa e negativa (-0,205). Portanto, quanto mais jovens sem atraso no ensino médio, menor a taxa de roubo. Embora a correlação seja baixa, essa direção é esperada pela teoria.
- **Taxa de atendimento escolar da população de 18 a 24 anos (T_FREQ18A24):** Esta variável, assim como a anterior, também representa frequência escolar. Ambas foram utilizadas para investigar qual teria maior correlação, e após qual teria melhor performance no modelo. A variável T_FREQ18A24 apresenta uma correlação linear baixa positiva (0,176). Esta foi a única variável de frequência

escolar a apresentar um resultado de correlação linear positivo (não intuitivo), embora muito baixo.

- **Taxa de atendimento escolar da população de 15 a 17 anos de idade (T_FREQ15A17):** Assim como a variável anterior, esta representa frequência escolar de indivíduos um pouco mais jovens. Os resultados permitem afirmar que a correlação linear entre T_FREQ15A17 e a taxa de roubo é baixíssima (-0,061), quase inexistente, ainda mais baixa que a T_FREQ18A24.
- **Taxa de frequência líquida ao ensino médio (T_FLMED):** Outra variável utilizada para sintetizar a frequência escolar, mas mais genérica que as anteriores. Apenas ao observar o gráfico de dispersão não é perceptível uma relação linear entre as variáveis. A correlação linear é baixa e negativa (-0,185), portanto, de forma muito sutil, quanto mais alta a taxa de frequência líquida ao ensino médio, menor a taxa de roubo.
- **Taxa de analfabetismo de indivíduos com 15 anos ou mais (T_ANALF15M):** Frequentemente utilizada na bibliografia, a taxa de analfabetismo usualmente apresenta relação positiva com a criminalidade (Bartz et al., 2018; W. Batella, 2008; Lima et al., 2005). A correlação linear entre a taxa de analfabetismo e a taxa de roubo no presente trabalho apresenta-se relativamente baixa e negativa (-0,271). Esta direção não é a esperada pela teoria: quanto maior a taxa de analfabetismo, menor a taxa de roubo. Mas, a relação entre as variáveis poderá não ser linear e variar no espaço.

A correlação linear alta diz respeito ao quanto uma variável aumenta (ou diminui) quando a outra variável aumenta. Quando essa relação é forte, pode ser um indicativo de que há uma relação explicativa entre as variáveis, porém não necessariamente. As variáveis independentes utilizadas neste estudo foram apontadas por outros autores como sendo potenciais variáveis explicativas da criminalidade (cf. Capítulo 2). Para investigar de fato se há relação de causa-efeito entre variáveis independentes e a variável dependente é necessário utilizar modelos

de regressão. Na próxima secção essas relações são exploradas em maior detalhe, quer através de modelos de regressão linear, quer através de modelos de regressão de Poisson.

5.4 Modelação da criminalidade associada a roubos

5.4.1 Modelos OLS

Após a pré-selecção das variáveis independentes, estas foram testadas múltiplas vezes, com combinações diferentes, em modelos OLS para investigar sua influência na taxa de roubos. O modelo que apresentou melhor poder explicativo sem evidência de multicolinearidade (Figura 14 e Figura 15) obteve um Adjusted R-Squared de 0,608. Ou seja, esse modelo explica 60% da variabilidade da taxa de roubo.

Figura 14 - Resultados do modelo OLS.

Summary of OLS Results - Model Variables								
Variable	Coefficient [a]	StdError	t-Statistic	Probability [b]	Robust_SE	Robust_t	Robust_Pr [b]	VIF [c]
Intercept	102,120695	4,390042	23,261895	0,000000*	4,345562	23,499996	0,000000*	-----
T_ATRASO_0	-9,156575	5,108186	-1,792530	0,073675	4,567247	-2,004835	0,045528*	1,353929
PREN20RICO	-19,635330	5,898566	-3,328831	0,000953*	5,403139	-3,634060	0,000321*	1,805326
RENOCUP	27,424961	7,422828	3,694678	0,000257*	6,756327	4,059152	0,000064*	2,858917
T_DES18M	27,567634	6,796433	4,056192	0,000065*	8,056488	3,421793	0,000689*	2,396761
T_NESTUDA_	-19,797233	7,059077	-2,804507	0,005243*	6,990684	-2,831945	0,004822*	2,585583
T_DENS	28,154128	6,687970	4,209667	0,000035*	7,078309	3,977522	0,000088*	2,320873
T_MULCHEFE	23,131135	5,489905	4,213394	0,000035*	5,087995	4,546218	0,000009*	1,563840
T_FUNDIN_T	-13,914447	6,742927	-2,063562	0,039580*	5,801197	-2,398547	0,016823*	2,359172
POP_ESTIMA	77,770144	5,011870	15,517192	0,000000*	17,734631	4,385214	0,000017*	1,303353

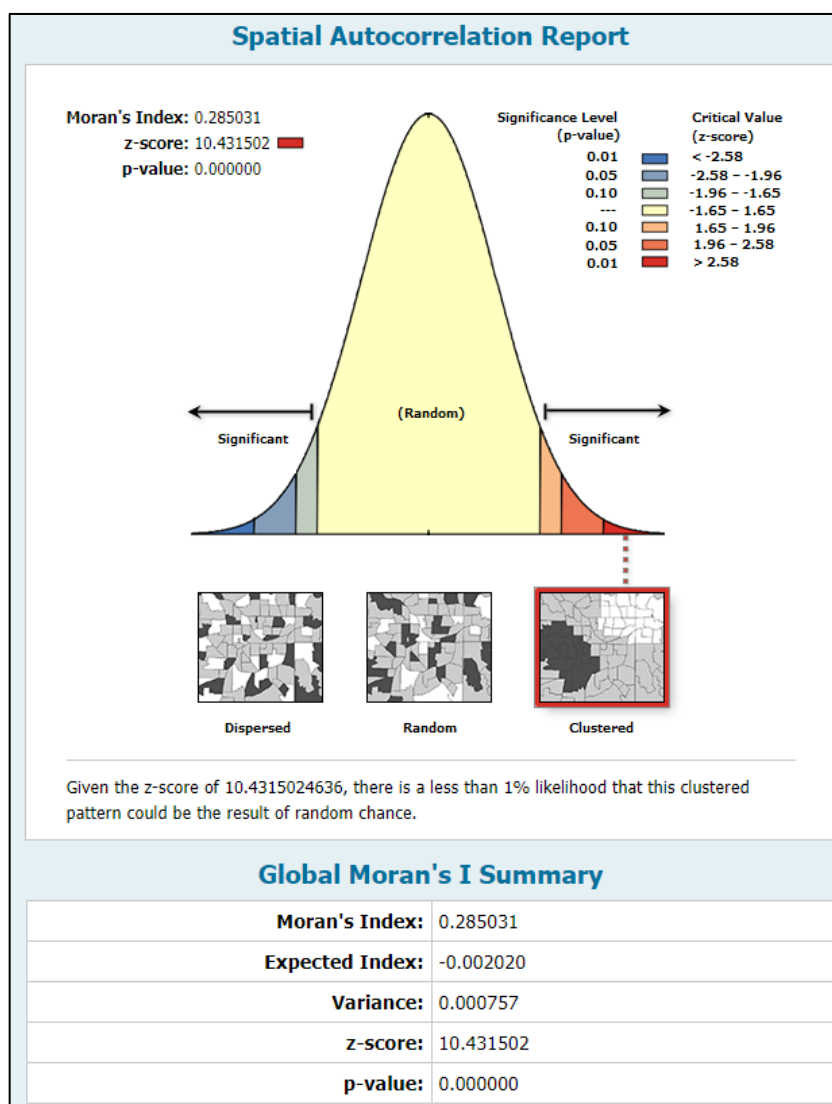
Figura 15 - Resultados e diagnósticos do modelo OLS.

OLS Diagnostics			
Input Features:	crime_2013_Conic_Zscore	Dependent Variable:	TAXA_ROUBO
Number of Observations:	496	Akaike's Information Criterion (AICc) [d]:	5965,996034
Multiple R-Squared [d]:	0,615664	Adjusted R-Squared [d]:	0,608547
Joint F-Statistic [e]:	86,502196	Prob(>F), (9,486) degrees of freedom:	0,000000*
Joint Wald Statistic [e]:	196,619617	Prob(> chi-squared), (9) degrees of freedom:	0,000000*
Koenker (BP) Statistic [f]:	90,987484	Prob(> chi-squared), (9) degrees of freedom:	0,000000*
Jarque-Bera Statistic [g]:	2304,856771	Prob(> chi-squared), (2) degrees of freedom:	0,000000*

Quanto aos diagnósticos, o teste Koenker permite concluir que pode haver heteroscedasticidade, i.e., não estacionaridade, das relações (p-value=0). Assim sendo, para avaliar a significância das variáveis é preciso examinar o Robust t-test, o qual permite concluir que todas as variáveis utilizadas neste modelo são significativas (p-value<0,05). O teste de Wald confirma a significância global do modelo (p-value=0), porém a estatística de Jarque-Bera infere que os resíduos não são distribuídos em uma normal (p-value=0). O teste de Moran Global I aplicado na sequência (Figura 16) revela que os resíduos têm autocorrelação espacial de agrupamento (p-value=0).

A não-normalidade dos resíduos quebra uma das premissas dos modelos de regressão linear, portanto o modelo pode estar enviesado. Todos os modelos testados apresentaram esse resultado.

Figura 16 - Relatório do teste de Moran Global I sobre os resíduos padronizados do modelo OLS.



De todas variáveis independentes testadas, estas (Figura 14) se revelaram estatisticamente significantes e foram as que, juntas, tiveram o maior poder explicativo para a taxa de roubo (60%). Essas variáveis representam as dimensões, que se mostraram mais importantes: **população** (POP_ESTIMA - população estimada), **educação** (T_ATRASO_0 - percentual de pessoas de 15 a 17 anos no ensino médio sem atraso), **renda** (PREN20RICO - percentual da renda apropriada pelos 20% mais ricos; e RENOCUP - rendimento médio dos ocupados - 18 anos ou mais), **ocupação** (T_DES18M - taxa de desocupação de pessoas com 18 anos ou mais; e T_NESTUDA_NTRAB_MMEIO - percentual de pessoas de 15 a 24 anos que não estudam nem trabalham e são vulneráveis à pobreza) e **ambiente familiar** (T_DENS - Percentual da população que vive em domicílios com densidade superior a 2 pessoas

por dormitório; T_MULCHEFEFIF04 - percentual de mães chefes de família sem fundamental completo e com filhos menores de 15 anos; e T_FUNDIN_TODOS - percentual de pessoas em domicílios em que ninguém tem fundamental completo).

Algumas dessas variáveis mostraram um coeficiente com sinal não intuitivo, nomeadamente PREN2ORICO, RENOCUP, T_NESTUDA_NTRAB_MMEIO e a T_FUNDIN_TODOS. Ao contrário do resultado no capítulo 5.3 e do esperado, a PREN2ORICO apresentou um coeficiente negativo na modelação global.

Intuitivamente se espera que o sinal da variável T_NESTUDA_NTRAB_MMEIO seja positivo. No capítulo 5.3 esta se mostrou não correlacionada com a taxa de roubo (0,009), porém no modelo OLS apresentou-se estatisticamente significativa e com o sinal negativo. O mesmo aconteceu com a variável T_FUNDIN_TODOS. Esperava-se um sinal positivo, porém ao interpretar o resultado negativo surge uma hipótese. Esta variável poderia representar, na verdade, a população rural, pois não é incomum que hajam domicílios onde ninguém tem o fundamental completo.

A variável RENOCUP apresentou um sinal positivo (não intuitivo), apesar de que, já argumentado no capítulo 5.3, o rendimento médio pode ter um sinal ambíguo, pois pode afetar diretamente os benefícios do crime e, também, pode mascarar a concentração de renda.

As variáveis restantes apresentaram um sinal de acordo com o esperado. Os maiores coeficientes são, respectivamente, das variáveis: população, T_DENS, T_DES18M e RENOCUP. Portanto a população, o ambiente domiciliar inapropriadamente denso, a taxa de desemprego e a renda média têm o maior peso no modelo.

Resultados não intuitivos podem ser consequência de uma interpretação errônea do que a variável representa. Ou ainda, conforme Cahill & Mulligan (2007), resultados não intuitivos no modelo global podem ocorrer por mascarar importantes variações locais nos parâmetros.

Embora o modelo tenha apresentado um bom poder explicativo, 40% da variabilidade da taxa de roubo ainda não é explicada e o modelo viola alguns dos seus pressupostos. O modelo pode estar sub-ótimo ou enviesado, pois os resíduos apresentaram autocorrelação espacial (i.e., não são independentes). Também, quando o teste Koenker é estatisticamente significativo no modelo OLS pode ser um indicativo

de que o relacionamento entre algumas ou todas as variáveis independentes e a variável dependente é não-estacionário. Este resultado já era esperado segundo a análise exploratória (secções 5.1.2 e 5.1.3). Neste contexto, é recomendável a utilização de modelos de regressão local para averiguar se o relacionamento entre as variáveis independentes e a variável dependente varia no espaço e se há algum aprimoramento do modelo.

5.4.2 Modelo GWR

Os resultados do modelo GWR (Tabela 12), construído com as mesmas variáveis do modelo OLS, revelam que o modelo apresentou uma melhora com a diminuição do AICc e com o aumento do R2Adjusted (de 60% no OLS para 73% no GWR), explicando 73% da variabilidade da taxa de roubo. Estes resultados reforçam a teoria de que os processos que influenciam as atividades criminosas são locais.

Tabela 12 - Resultados e diagnósticos do modelo GWR.

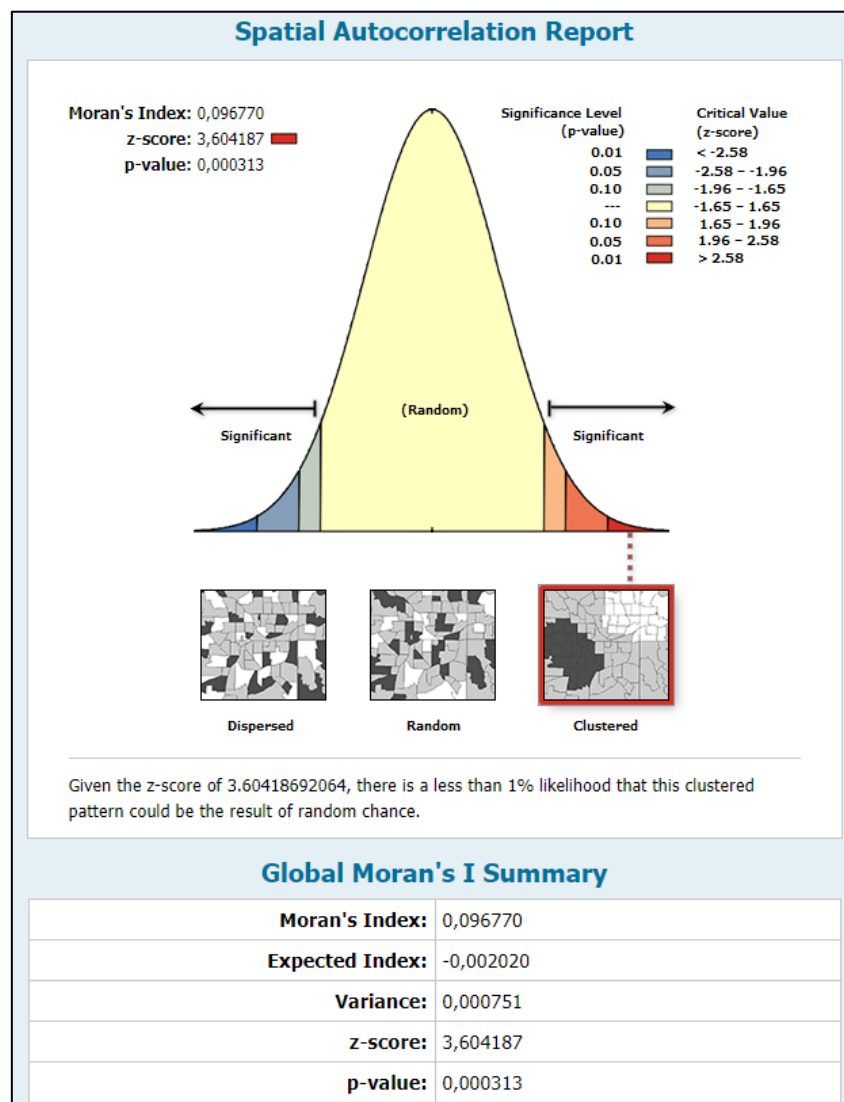
OBJECTID *	VARNAME	VARIABLE	DEFINITION
1	Neighbors	191	
2	ResidualSquares	2667436,076805	
3	EffectiveNumber	84,341695	
4	Sigma	80,496791	
5	AICc	5819,039121	
6	R2	0,779327	
7	R2Adjusted	0,734651	

O Condition Number para cada município varia entre 3,44 e 7,46 (< 30), sendo assim, o modelo não apresenta multicolinearidade. Os R-Squared locais variam entre 0,34 e 0,76. Este resultado indica que o modelo se ajusta bem em alguns municípios, porém em outros talvez faltem variáveis explicativas importantes.

Comparado ao modelo OLS, o modelo GWR apresentou uma melhora também na distribuição dos resíduos (Figura 17), tanto o índice de Moran como o z-score diminuíram consideravelmente. Este fato é interessante e serve de evidência de que o GWR pode resolver (ou amenizar) o problema dos termos de erro espacialmente autocorrelacionados, conforme Fotheringham, Brunson, & Charlton (2002, p. 117) haviam afirmado.

Embora a melhora considerável no modelo, os resíduos continuaram a apresentar autocorrelação espacial (não independência). Neste caso, o modelo não é totalmente confiável, podendo estar enviesado, subestimando ou superestimando o valor da variável dependente. No contexto deste estudo, isto pode se dar porque há variáveis explicativas importantes que não foram incluídas no modelo e/ou porque a relação (local) entre as variáveis explicativas e a variável dependente não é linear. Por outro lado, os dados que estamos trabalhando se tratam de taxas baseadas em dados de contagem. Neste caso se justificou a construção de modelos de regressão de Poisson, pois são mais adequados a este tipo de dados e foram também utilizados na modelação da criminalidade (cf. revisão de literatura).

Figura 17 - Relatório do teste de Moran Global I sobre os resíduos padronizados do modelo GWR.



5.4.3 Modelos Poisson

Efetivamente, as taxas de roubo são baseadas na contagem do número de roubos. É comum que cidades menores, com poucos habitantes, obtenham poucos (ou nenhum) registros de roubos por ano. Para o ano de 2013 no RS, muitas cidades, como visto anteriormente (seção 5.1 Análise exploratória da criminalidade associada a roubos), marcaram 0 roubos ou tem a taxa de roubo muito próxima a 0. Isso pode gerar problemas na regressão linear, como a obtenção de estimativas OLS negativas.

Ao comparar os resultados do Poisson Global (Figura 18) e do GWPR (Figura 19) observa-se que o AICc é menor no modelo local (GWPR). Isso permite concluir que o modelo local tem uma performance melhor do que o global. Além disso a percentagem de *deviance* explicada passa de 58% no modelo global para 68% no modelo local. Infelizmente, não é possível comparar o AICc de modelos de Poisson com os de modelos lineares, como visto anteriormente (seção 4.5 Modelos de Poisson).

Figura 18 - Resultado Poisson Global.

```
*****
Global regression result
*****
< Diagnostic information >
Number of parameters:          10
Deviance:                     32138.838754
Classic AIC:                   32158.838754
AICc:                          32159.292362
BIC/MDL:                       32200.904513
Percent deviance explained     0.580701
```


Figura 19 - Resultado Poisson Local (GWPR).

```

*****
GWR (Geographically weighted regression) result
*****

Bandwidth and geographic ranges
Bandwidth size: 60.836849
Coordinate      Min      Max      Range
-----
X-coord      -320851.605571  406384.180460  727235.786031
Y-coord      -3870786.231960 -3152782.877390  718003.354570

Diagnostic information
Effective number of parameters (model: trace(S)): 40.394574
Effective number of parameters (variance: trace(S'WSW^-1)): 26.696352
Degree of freedom (model: n - trace(S)): 455.605426
Degree of freedom (residual: n - 2trace(S) + trace(S'WSW^-1)): 441.907203
Deviance: 24442.637324
Classic AIC: 24523.426472
AICc: 24530.782813
BIC/MDL: 24693.349316
Percent deviance explained 0.681109

```

Quanto ao modelo semi-paramétrico (S-GWPR), o relatório da ferramenta Local-To-Global (Figura 20) permite concluir que não há aperfeiçoamento no modelo com a afixação de coeficientes globais. Portanto o modelo GWPR continua sendo o mais apropriado para os dados. Os resultados do modelo GWPR são apresentados e discutidos detalhadamente em seguida.

Figura 20 - Resultado da ferramenta Local-To-Global.

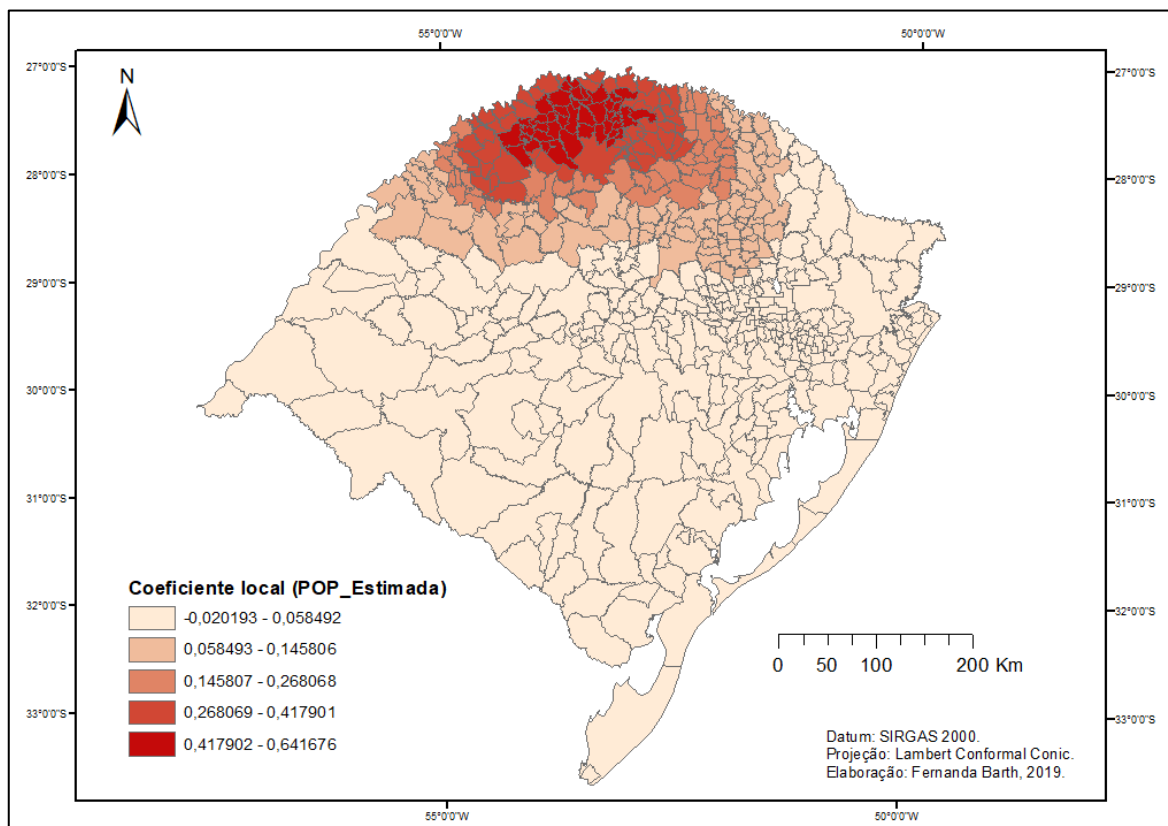
The summary of the L -> G variable selection model	AICc
GWR model before L -> G selection	24530.782813
GWR model after L -> G selection	24530.782813
Improvement	0.000000

A interpretação dos valores dos coeficientes não é tão intuitiva nos modelos de Poisson como no caso dos modelos lineares. A regressão de Poisson modela o log do valor esperado da variável dependente como uma função das variáveis explicativas. Se a variável explicativa aumentar uma unidade, assumindo que as outras variáveis independentes no modelo são mantidas constantes, espera-se que o log do valor esperado da contagem (ou taxa) seja o valor do respectivo coeficiente de

regressão. O objetivo do modelo obtido não é fazer previsão, mas sim investigar como varia espacialmente a relação entre as variáveis explicativas e a criminalidade. Portanto, a análise dos coeficientes locais irá focar-se na variabilidade espacial da sua magnitude e sinal. Tal como nos modelos OLS, o sinal do coeficiente indica a direção da relação. Um sinal positivo indica que ambas as variáveis tendem a aumentar ou diminuir em conjunto, e um sinal negativo significa que elas se movem em direções opostas.

O mapa do coeficiente da população (Figura 21) ilustra bem a não-estacionariedade. Enquanto a variável tem um coeficiente mais alto (e positivo) na região norte, no resto do estado a variável tem um valor menor e até mesmo negativo. Portanto, é possível concluir que o fator população é mais importante na região norte, onde quanto maior a população maior a taxa de roubo. Já no resto do estado a população não tem um papel tão importante, e a relação é negativa.

Figura 21 - Mapa do coeficiente da variável população estimada por município no RS.

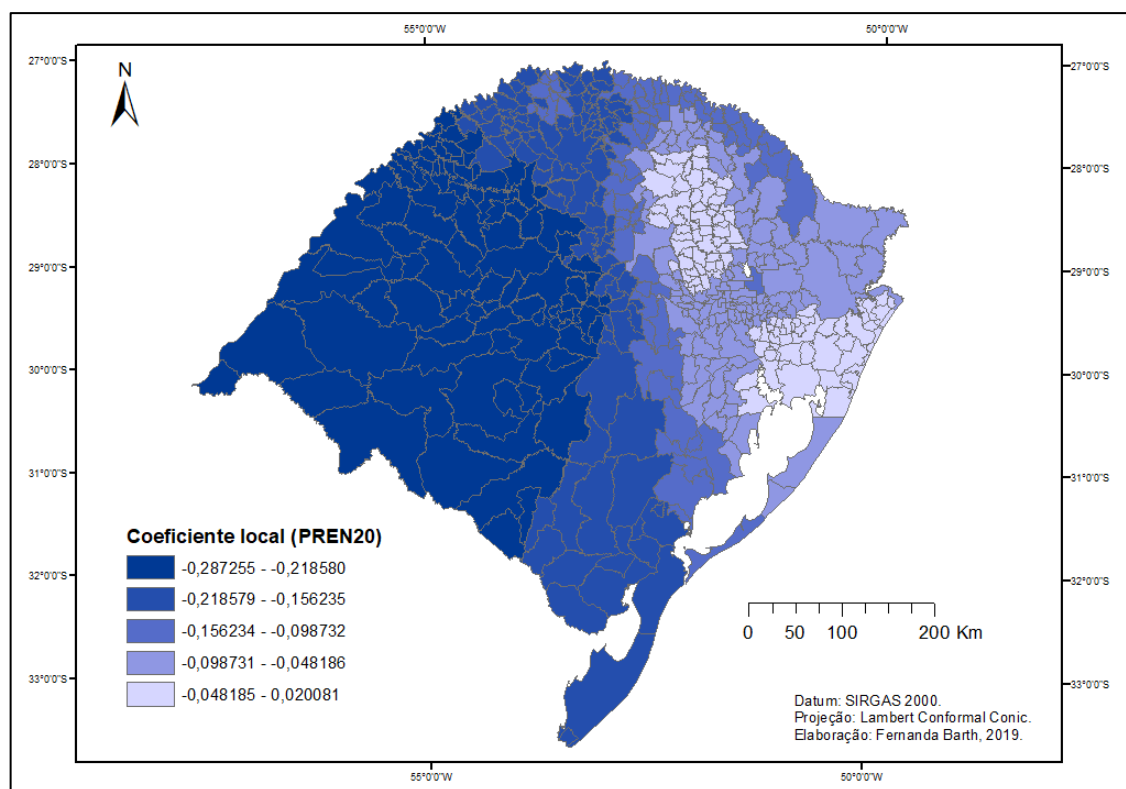


Para modelos locais GWR não existem testes estatísticos para avaliar a significância dos coeficientes. Por este motivo, usam-se os erros padrão (ou os *pseudo-t values*) para examinar a fiabilidade dos coeficientes. Se os *standard errors* (erros padrão) forem mais elevados em determinada região, indica que os coeficientes têm menos precisão nesta região, e poderão não ser fiáveis. Os mapas dos erros padrão encontram-se em anexo.

O mapa dos erros padrão para todos os coeficientes, indicam que na região norte e, às vezes, na região metropolitana de Porto Alegre os erros padrão são maiores. Pode ser porque os coeficientes não são estatisticamente significativos nestas regiões (i.e. nessas regiões as variáveis não tem poder explicativo).

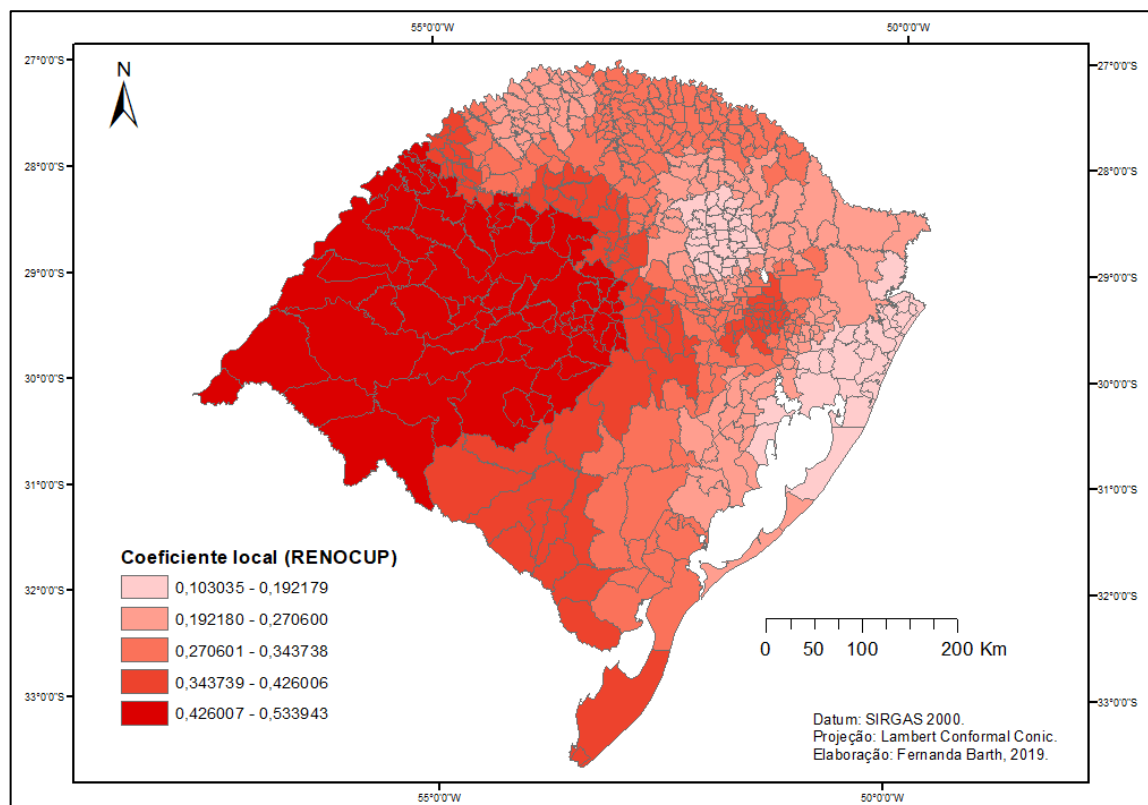
A variável PREN20RICO (percentual da renda apropriada pelos 20% mais ricos) tem um coeficiente negativo em quase toda a área de estudo (Figura 22). Porém nota-se uma diferença marcante entre as regiões leste e oeste. A região oeste é mais afetada pela variável que representa desigualdade de renda. Já no leste a variável tem um menor poder explicativo sobre os roubos.

Figura 22 - Mapa do coeficiente da variável PREN20RICO por município no RS.



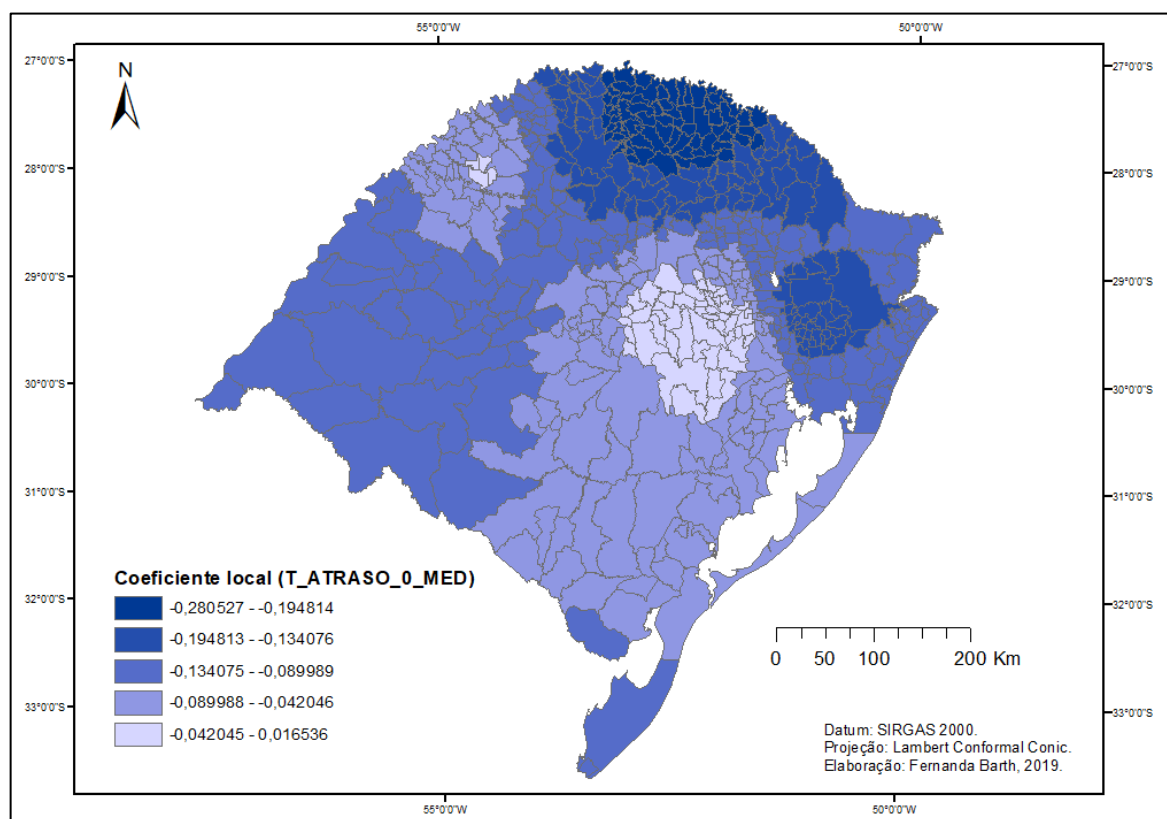
O coeficiente de RENOCUP (rendimento médio dos ocupados - 18 anos ou mais) muda bastante de região para região (Figura 23), embora mantenha o sinal positivo. Enquanto na região leste, sobretudo na região do litoral norte e na região da serra gaúcha, o coeficiente é mais baixo, na região oeste o coeficiente é bastante alto, significando que o rendimento médio dos ocupados é um fator muito relevante para a taxa de roubos na região.

Figura 23 - Mapa do coeficiente da variável RENOCUP por município no RS.



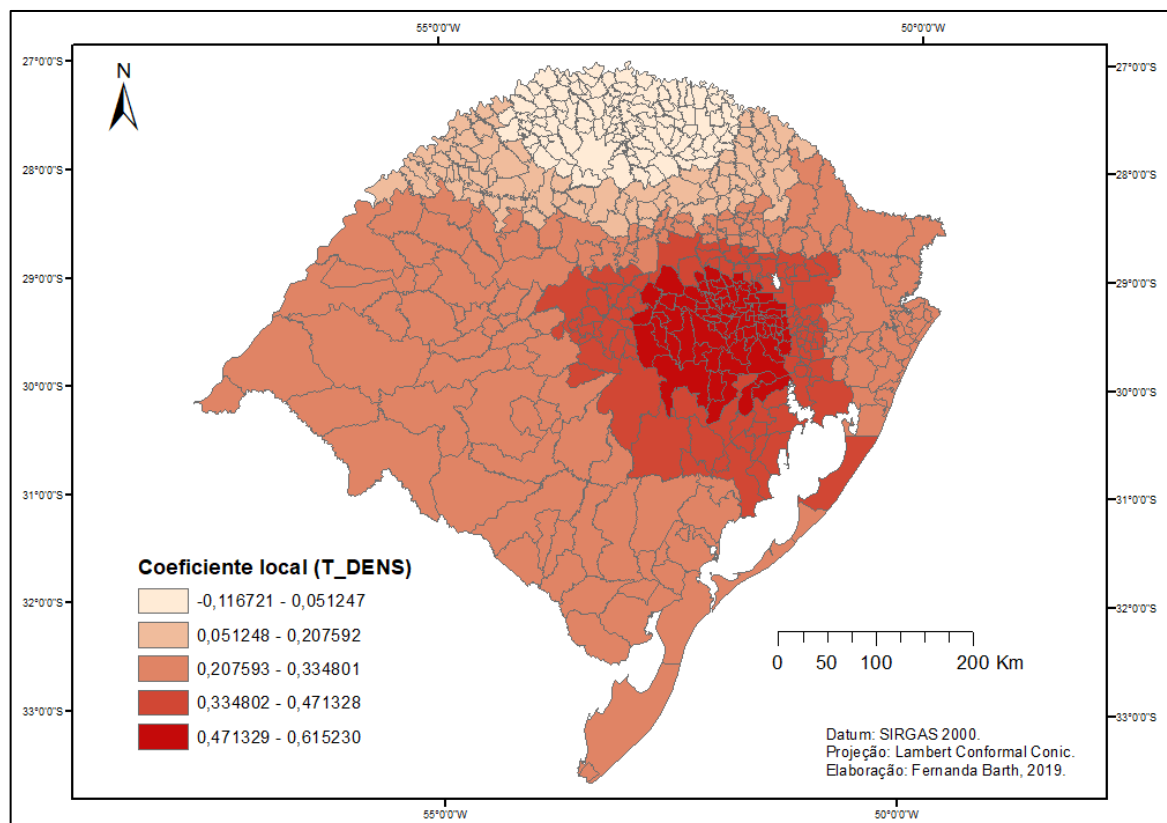
Quanto ao coeficiente da variável T_ATRASO_0_MED (percentual de pessoas de 15 a 17 anos no ensino médio sem atraso) por município (Figura 24), este mantém-se com sinal negativo por quase todo o estado. Na região central do estado este é mais baixo e na região norte é mais alto. Ou seja, na região norte a variável tem maior poder explicativo sobre os roubos. Portanto, campanhas para reduzir o atraso escolar teriam mais efeito na redução da criminalidade para as regiões norte, nordeste, região metropolitana de Porto Alegre e região oeste do que para a região centro-sul.

Figura 24 - Mapa do coeficiente da variável T_ATRASO_0_MED por município no RS.



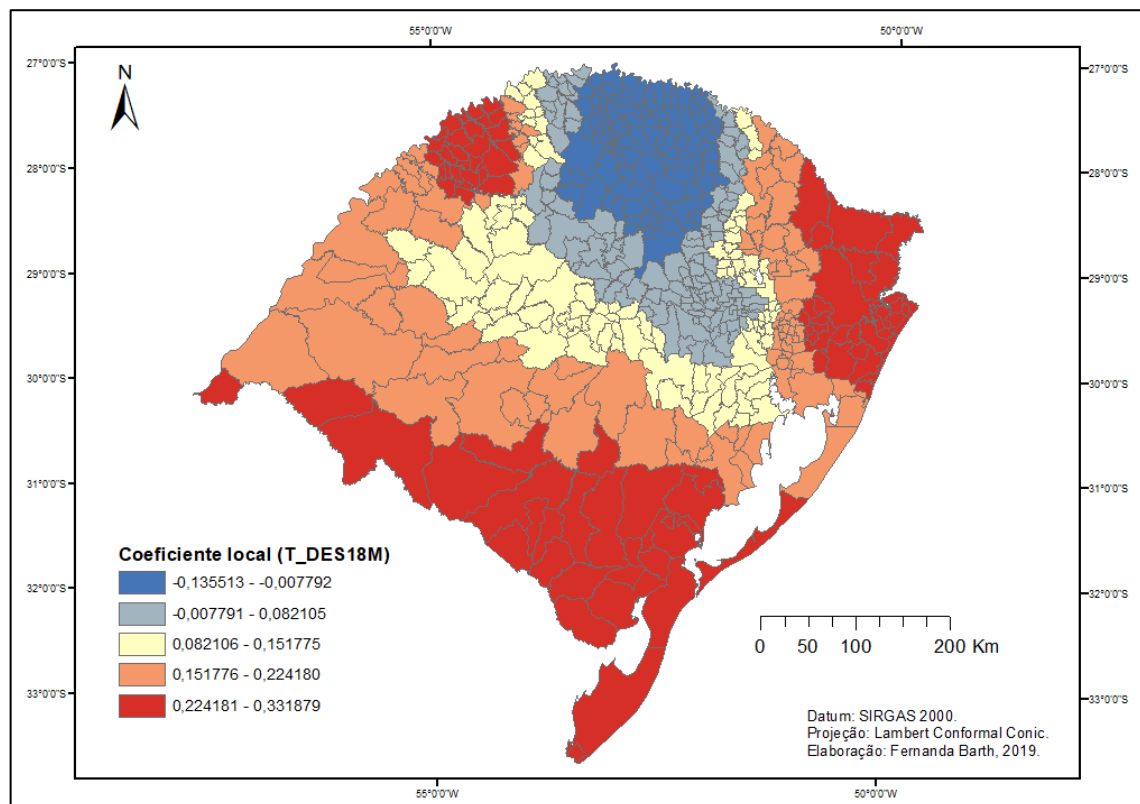
O coeficiente da variável T_DENS (Percentual da população que vive em domicílios com densidade superior a 2 pessoas por dormitório) varia bastante seu valor entre os municípios (Figura 25). O sinal da variável é positivo em grande parte da área de estudo. Enquanto na região norte essa variável não tem uma grande influência sobre a taxa de roubo, na região centro oriental, incluindo a região metropolitana de Porto Alegre, a variável tem uma forte influência.

Figura 25 - Mapa do coeficiente da variável T_DENS por município no RS.



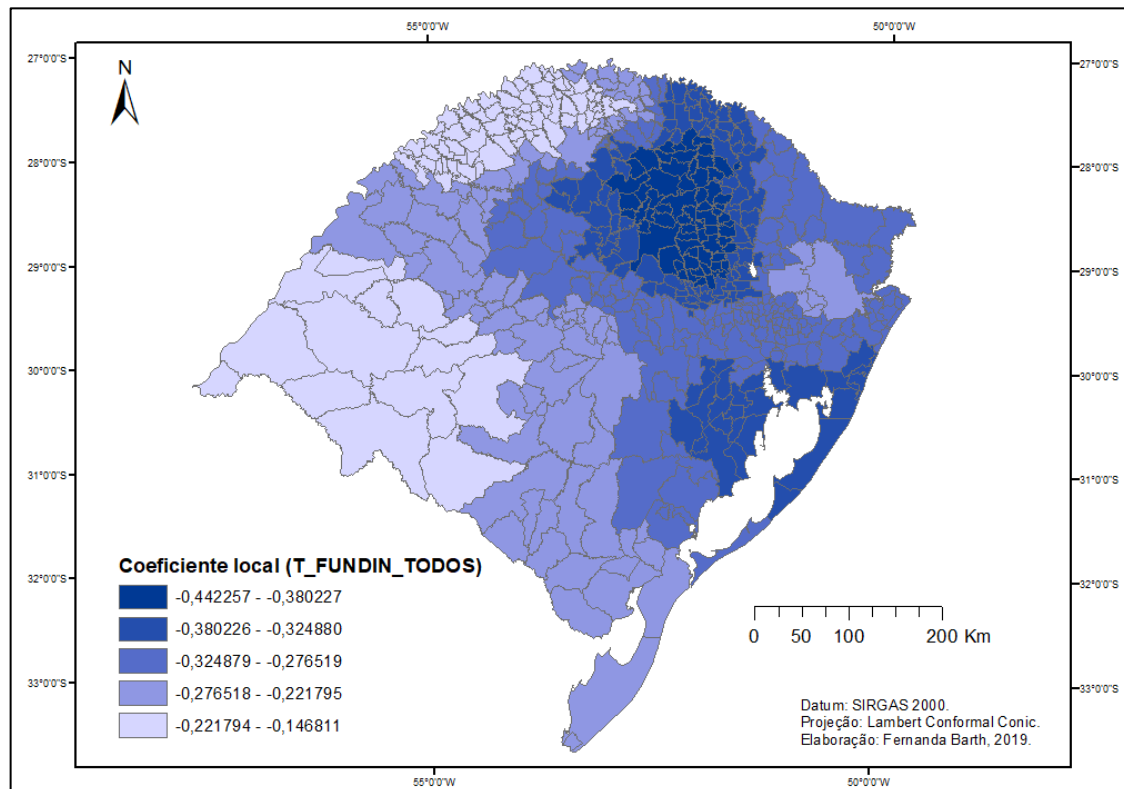
O mapa do coeficiente da variável T_DES18M (taxa de desocupação de pessoas com 18 anos ou mais) por município (Figura 26) permite concluir que o valor do coeficiente varia de sinal positivo para sinal negativo no estado. Na região norte a variável T_DES18M tem um sinal negativo não intuitivo, quando esta aumenta, a taxa de roubo diminui. Porém nesta região, onde o sinal negativo se apresenta, os erros padrão são maiores, assim, o coeficiente não é fiável. Já no restante do estado a variável tem uma influência maior e positiva sobre as taxas de roubo, quanto maior a taxa de desocupação (18+), maior a taxa de roubo.

Figura 26 - Mapa do coeficiente da variável T_DES18M por município no RS.



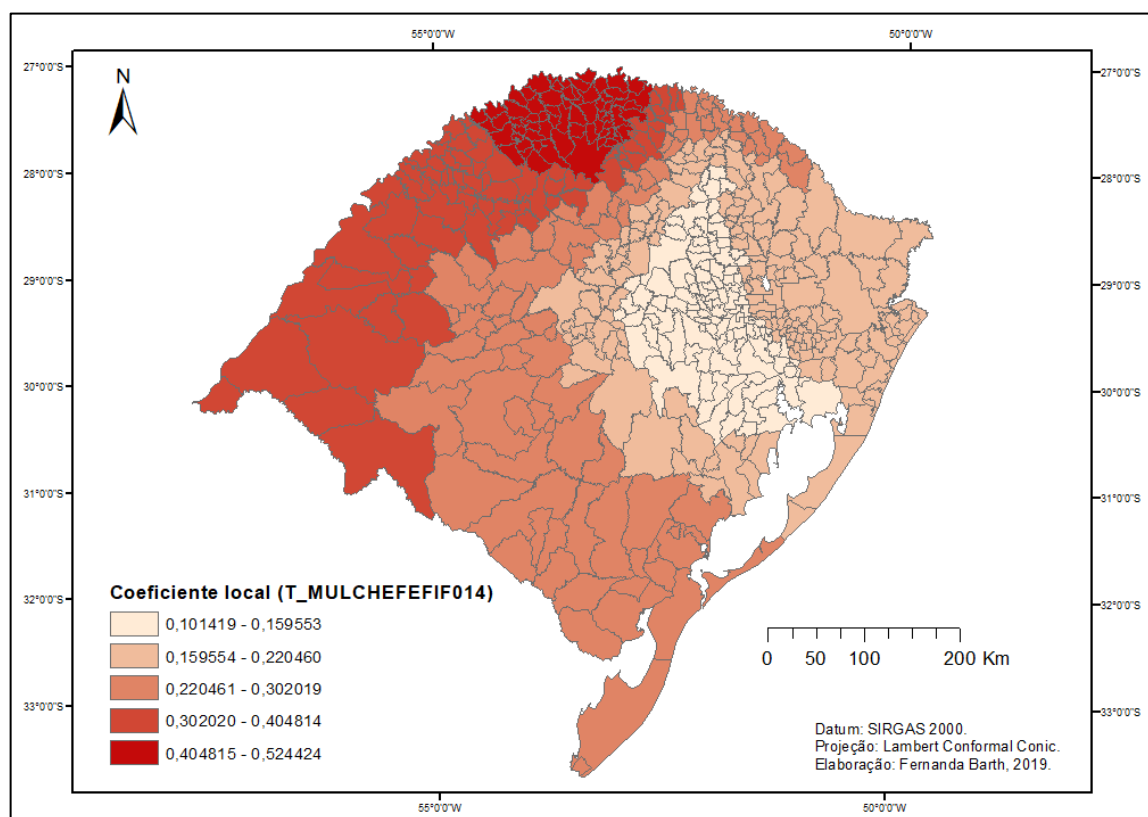
O coeficiente da variável T_FUNDIN_TODOS (percentual de pessoas em domicílios em que ninguém tem fundamental completo) apresenta um sinal negativo, não intuitivo, assim como no modelo OLS, por toda a área de estudo (Figura 27). Porém, enquanto na região oeste e norte a influência é menor, nas regiões centro-nordeste e sudeste, a influência da variável é maior.

Figura 27 - Mapa do coeficiente da variável T_FUNDIN_TODOS por município no RS.



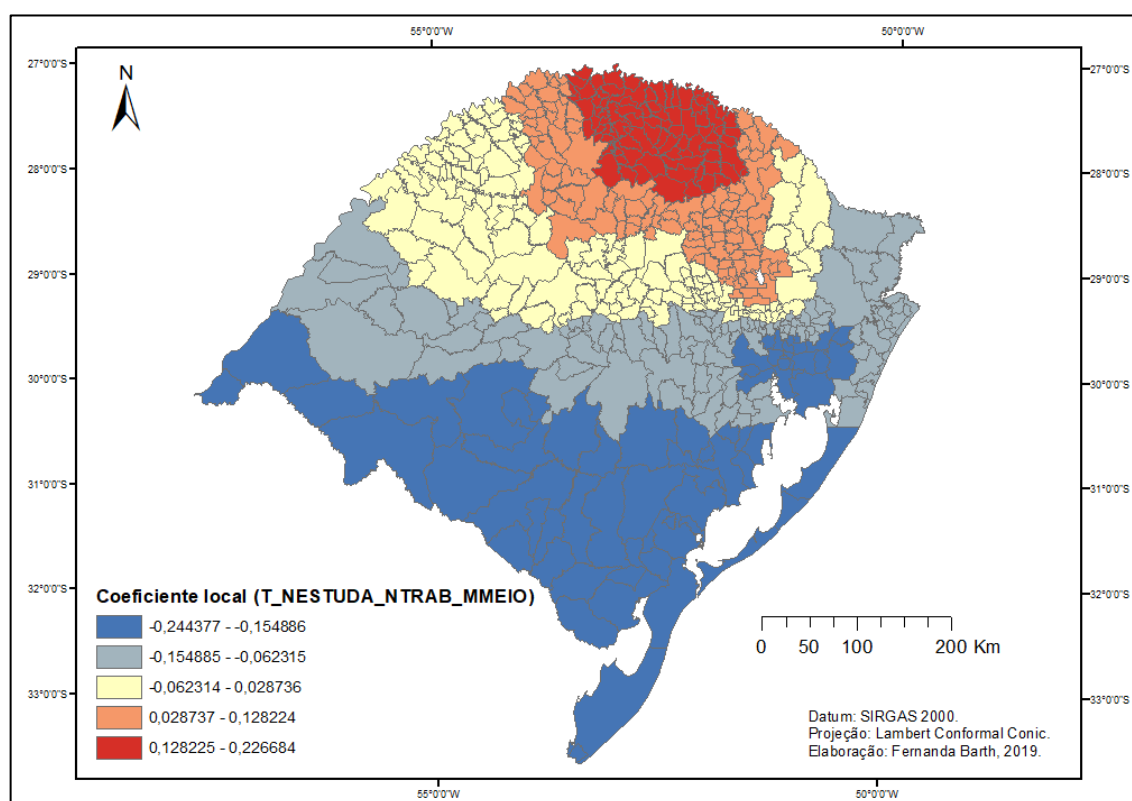
O coeficiente da variável T_MULCHEFEFIF014 (percentual de mães chefes de família sem fundamental completo e com filhos menores de 15 anos) se mantém positivo por todo o estado (Figura 28). No entanto, existem diferenças quanto ao potencial explicativo. Enquanto na região centro-leste (incluindo a capital) a variável tem menor potencial explicativo, nas regiões de norte à oeste o percentual de mães chefes de família é mais importante para explicar suas respectivas taxas de roubos. Este resultado na região norte pode não ser fiável, pois o erro padrão na região é alto.

Figura 28 - Mapa do coeficiente da variável T_MULCHEFEFIF014 por município no RS.



O mapa do coeficiente da variável T_NESTUDA_NTRAB_MMEIO (percentual de pessoas de 15 a 24 anos que não estudam nem trabalham e são vulneráveis à pobreza) também demonstra bem a não-estacionariedade (Figura 29), pois em algumas regiões o sinal é negativo e em outras o sinal é positivo. O efeito de norte a sul é claro, ao norte a variável tem um sinal positivo e este vai diminuindo e se tornando negativo no sentido sul. O sinal positivo para esta variável não é esperado, além disto o erro padrão para a região norte é alto, portanto pode significar que a variável não tem potencial explicativo na região.

Figura 29 - Mapa do coeficiente da variável T_NESTUDA_NTRAB_MMEIO por município no RS.



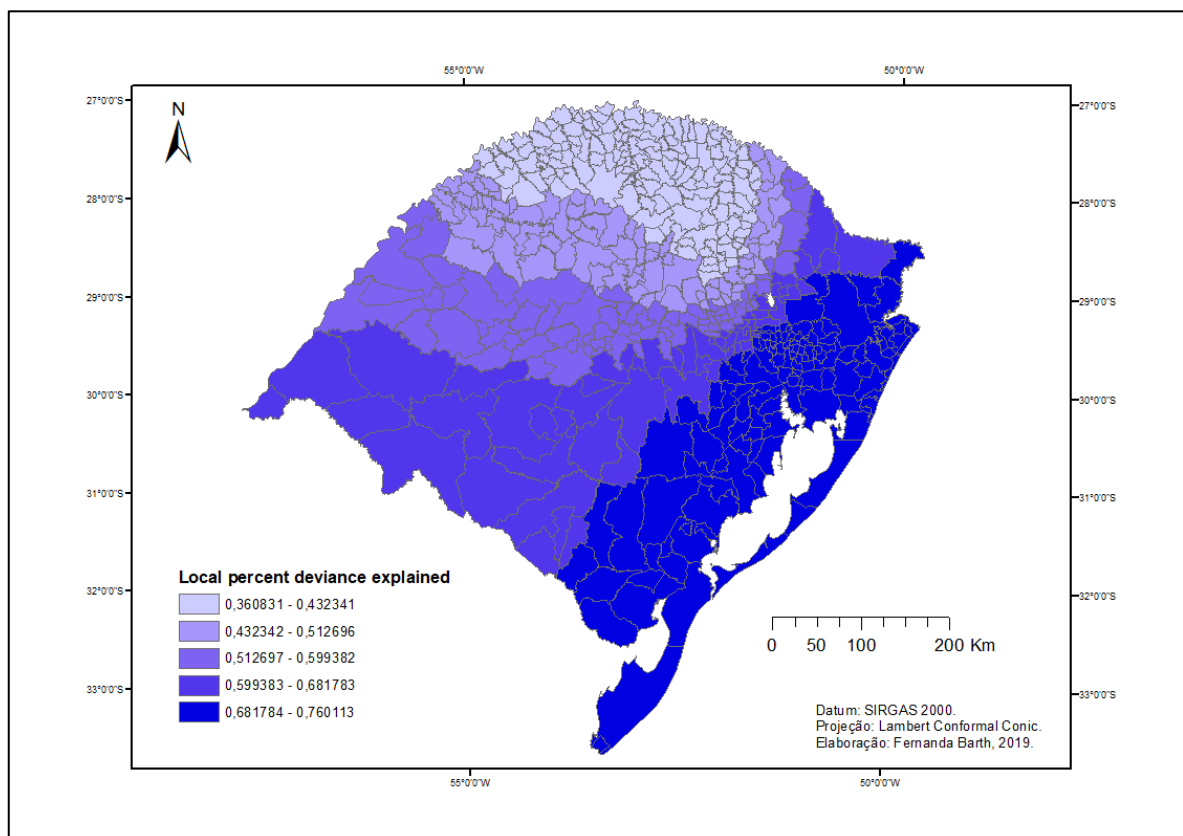
Estas variações de magnitude e sinal da mesma variável para diferentes regiões revela o quão importante é considerar a não estacionariedade das relações. Um modelo global mascara estas diferenças.

O mapa de std residual está apresentado em anexo (Figura XX dos Anexos), sendo a interpretação da seguinte forma: se está em tons de azul no mapa é porque o resíduo é negativo e isso significa que o valor real da taxa de roubo neste município é menor do que o previsto pelo modelo e vice-versa. As regiões com os valores mais

elevados em termos absolutos indicam as áreas onde o modelo tem pior ajustamento local aos dados, nomeadamente em Pelotas, Candiota, Charqueadas, Eldorado do Sul, Porto Alegre, Nova Santa Rita, Capivari do Sul, Capela de Santana, Montenegro, Esteio, Sapucaia do Sul, Novo Hamburgo, Campo Bom, Capão da Canoa, Mato Castelhano e Gentil.

Por fim, o mapa do percentual da *deviance* explicada permite concluir que não apenas o poder explicativo das variáveis muda de município para município, como também o poder explicativo do próprio modelo. O modelo GWPR explica melhor as taxas de roubo nos municípios do sul a nordeste do estado (Figura 30). Já ao norte o potencial explicativo do modelo é menor, sendo provável que falem variáveis explicativas importantes para esta região no modelo.

Figura 30- Mapa do percentual de *deviance* explicada pelo modelo por município no RS.



6. CONCLUSÃO

Este estudo teve como principais objetivos a análise dos padrões da criminalidade no Rio Grande do Sul, em particular das taxas de roubo e de homicídio, bem como a identificação dos fatores que podem causar os padrões observados em 2013.

A média da taxa de roubos por 100 mil habitantes no Rio Grande do Sul aumentou em torno de 52% entre 2013 e 2017, enquanto que a taxa de homicídios aumentou aproximadamente 17% neste período. Quanto as taxas de homicídio, a análise exploratória permitiu concluir que não há um padrão definido na distribuição das mesmas no Rio Grande do Sul para o ano de 2013. A distribuição é aleatória e o número de ocorrências é baixo ou nulo para quase todos os municípios, com quase nenhuma presença de Hot Spots. Na análise exploratória obteve-se resultados muito diferentes entre a taxa de homicídio e a taxa de roubo. O crime de homicídio aparenta ter outras motivações e causas que não resultam nem são influenciadas a partir de um processo regional. Perante estes resultados, não se justificou a análise da taxa de homicídio através de modelos de regressão espacial.

Este trabalho possibilitou a identificação de importantes fatores relacionados a taxa de roubos no Rio Grande do Sul. As variáveis explicativas da taxa de roubo no estado representam as dimensões de população, renda, ocupação, educação e ambiente familiar. Estes resultados reforçam a teoria de Oliveira (2008) onde afirma que os fatores que levam um indivíduo a cometer um crime, mais especificamente de roubo, não podem ser explicados apenas pela teoria econômica, mas também pelos microssistemas de desenvolvimento do indivíduo que influenciam no desenvolvimento da moral.

Além disso, os processos que levam um indivíduo a cometer roubo, transcendem o micro ambiente familiar e são diferentes de local para local (o macro ambiente). Ou seja, estes processos causais não são estacionários, o que demonstra também que estes são processos complexos.

A modelação local (GWR/GWPR) proporciona um aprofundamento neste tema, pois permite explorar detalhadamente, de local a local, quais fatores tem maior

poder explicativo. Este trabalho reforça a afirmação de Vilalta e Muggah (2016, p. 7) de que quando há heterogeneidade espacial, um modelo global não se ajusta adequadamente aos dados, como o OLS ou o modelo de Poisson global, portanto os resultados são no mínimo sub-ótimos e muitas vezes tendenciosos, pois relações locais podem se anular mutuamente no cálculo de estimativas globais. Portanto, o uso de modelos locais é extremamente recomendável em estudos sobre criminalidade, sobretudo relacionados a roubo.

Entre os modelos lineares OLS e GWR, se constatou que o GWR teve uma melhor performance. No entanto, ambos modelos apresentaram resíduos espacialmente autocorrelacionados. A resposta para este problema pode estar na natureza dos dados da taxa de roubo, que tem como base dados de contagem. Este problema é decorrente e já se tem discutido a sua solução com o uso de modelos de Poisson (Chen et al., 2017; Osgood, 2000; Vilalta & Fondevila, 2019).

Entre os modelos de Poisson global, GWPR e S-GWPR, o modelo GWPR apresentou o melhor ajustamento. Estes resultados reforçam mais uma vez que as relações entre as variáveis explicativas e a variável dependente são não-estacionárias, tal como os resultados da análise exploratória também evidenciaram.

Os resultados levam a concluir que existe tanto heterogeneidade espacial quanto autocorrelação espacial na taxa de roubo. Municípios próximos possuem valores similares da taxa de roubo, enquanto municípios distantes possuem valores diferentes e, às vezes, causas diferentes. Ou seja, o crime de roubo é um processo regional e, portanto, sua solução também. Políticas para redução da sua ocorrência devem ser direcionadas e levar em conta os reais fatores causais da taxa de roubo naquela região.

6.1 Limitações do estudo

Este trabalho teve algumas limitações, como a disponibilidade de variáveis socioeconômicas dos municípios para os anos mais recentes (e.g. 2017), pois esse tipo de variável é disponibilizada majoritariamente através dos censos, e estes são realizados a cada 10 anos, limitando o estudo para os anos mais recentes. A análise foi

feita para o ano de 2013 por ser mais próximo de 2010, assumindo que essas variáveis não tiveram grandes alterações nestes 3 anos.

O eventual problema de sobre dispersão do modelo GWPR poderia ser corrigido através da inclusão de um parâmetro de escala no modelo (i.e., estimando um modelo quasi-poisson), o qual permite obter estatísticas ajustadas (erros padrão, *deviance residuals*, etc.). Contudo, ficou fora do âmbito do trabalho uma vez que o software GWR4 não tem esta opção e a utilização do package ‘spgwr’ (Bivand, Yu, Nakaya, Garcia-Lopez, & Bivand, n.d.), ou do package ‘GWmodel’ do software R (Gollini, Lu, Charlton, Brunsdon, & Harris, 2015; Lu, Harris, Charlton, & Brunsdon, 2014), não foi possível por limitações de tempo.

6.2 Trabalhos futuros

Para trabalhos futuros é sugerida a utilização do modelo de regressão Binomial Negativa. Este modelo é frequentemente usado para resolver o problema de sobre-dispersão. Contudo, Berk & MacDonald (2008) defendem que tal pode conduzir a um pior modelo porque, na prática, muitas vezes trata-se de um problema de aparente sobre-dispersão (apparent overdispersion).

Outras possibilidades de análise para expandir o estudo poderão envolver a utilização de metodologias de Geospatial Data Mining, sobretudo através de clustering, para investigar padrões e relações espaciais na base de dados de interesse.

REFERÊNCIAS

- Anselin, L. (1995). Local Indicators of Spatial Association - LISA. *Geographical Analysis*, 27(2), 93–115.
- Anselin, L. (2004). Appendix C – Ordinary Least Squares and Poisson Regression Models. In *CrimeStat III: A Spatial Statistics Program for the Analysis of Crime Incident Locations*. Houston, TX & Washington, DC: Ned Levine.
- Bai, J., & Ng, S. (2005). Tests for skewness, kurtosis, and normality for time series data. *Journal of Business and Economic Statistics*, 23(1), 49–60.
<https://doi.org/10.1198/073500104000000271>
- Bartz, M., Quartieri, E., & Menezes, G. (2018). Criminalidade no Rio Grande do Sul: Uma Análise Econométrica para os Coredes no ano de 2010*. *Revista Brasileira de Estudos Regionais e Urbanos (RBERU)*, 12(1), 110–128.
- Batella, W. (2008). *ANÁLISE ESPACIAL DOS CONDICIONANTES DA CRIMINALIDADE VIOLENTA NO ESTADO DE MINAS GERAIS – 2005: Contribuições da Geografia do Crime* (Pontifícia Universidade Católica de Minas Gerais). Retrieved from http://www.sistemas.pucminas.br/BDP/SilverStream/Pages/pg_Consltem.html
- Batella, W. B., & Diniz, A. M. A. (2010). Análise Espacial dos Condicionantes da Criminalidade Violenta no Estado de Minas Gerais. *Revista Sociedade & Natureza*, 22(1), 151–163. Retrieved from <http://www.seer.ufu.br/index.php/sociedadennatureza/article/view/9747>
- Berk, R., & MacDonald, J. M. (2008). Overdispersion and poisson regression. *Journal of Quantitative Criminology*, 24(3), 269–284. <https://doi.org/10.1007/s10940-008-9048-4>
- Bernasco, W., & Elffers, H. (2010). Statistical Analysis of Spatial Crime Data. In *Handbook of Quantitative Criminology* (pp. 699–724). <https://doi.org/10.1007/978-0-387-77650-7>
- Bivand, R., Yu, D., Nakaya, T., Garcia-Lopez, M. A., & Bivand, M. R. (n.d.). Package ‘spgwr’. R software package. Retrieved from <https://cran.r-project.org/package=spgwr>
- Brunsdon, C., Fotheringham, A. S., & Charlton, M. E. (1996). Geographically Weighted Regression: A Method for Exploring Spatial Nonstationarity. *Geographical*

- Analysis*, 28(4), 281–298. <https://doi.org/10.1111/j.1538-4632.1996.tb00936.x>
- Cahill, M., & Mulligan, G. (2007). Using geographically weighted regression to explore local crime patterns. *Social Science Computer Review*, 25(2), 174–193. <https://doi.org/10.1177/0894439307298925>
- Carrets, F. D., Oliveira, J. De, & Menezes, G. (2018). A criminalidade no Rio Grande do Sul: uma análise espacial para anos de 2005, 2010 e 2015. *Perspectiva Econômica*, 14(1), 33–46. <https://doi.org/10.4013/pe.2018.141.03>
- Cerqueira, D., Lima, R., Bueno, S., Neme, C., Ferreira, H., Coelho, D., ... Merian, F. (2018). *Atlas da Violência 2018*. Retrieved from http://www.ipea.gov.br/portal/images/stories/PDFs/relatorio_institucional/180604_atlas_da_violencia_2018.pdf
- Chen, J., Liu, L., Zhou, S., Xiao, L., Song, G., & Ren, F. (2017). Modeling spatial effect in residential burglary: A case study from ZG City, China. *ISPRS International Journal of Geo-Information*, 6(5). <https://doi.org/10.3390/ijgi6050138>
- Fotheringham, A. S., Brunson, C., & Charlton, M. (2002). *Geographically Weighted Regression: the analysis of spatially varying relationships* (Wiley, Ed.). Chichester, UK.
- Givisiez, G., & Oliveira, E. (2013). Privacidade intradomiciliar : um estudo sobre as necessidades de ampliações em residências*. *Revista Brasileira de Estudos de População*, 30(1), 199–223.
- Gollini, I., Lu, B., Charlton, M., Brunson, C., & Harris, P. (2015). GWmodel : an R package for exploring spatial heterogeneity. *Journal of Statistical Software*, 63(17), 1–50. <https://doi.org/10.1080/10095020.2014.917453>
- Gomes, C. E., Evangelista, T. F., Lima, R. L., & Parré, J. L. (2017). Determinantes do crime nos municípios de Minas Gerais e seus possíveis spillovers espaciais. *Revista Economia Ensaios*, 31(2), 201–228. <https://doi.org/10.14393/ree-v31n2a2017-9>
- Hurvich, C., Simonoff, J., & Tsai, C. (1998). Smoothing parameter selection in nonparametric regression using an improved Akaike information criterion. *Journal of the Royal Statistical Society: Series B*, 60(2), 271–293. <https://doi.org/10.1109/ISSPA.2005.1581081>
- IBGE. (2019). Panorama Rio Grande do Sul. Retrieved August 5, 2019, from

- <https://cidades.ibge.gov.br/brasil/rs/panorama>
- IPEA. (2019). Sobre o Atlas da Violência. Retrieved July 20, 2019, from <http://www.ipea.gov.br/atlasviolencia/quem>
- Koenker, R., Bassett, G., & Jr. (1982). Robust Tests for Heteroscedasticity Based on Regression Quantiles. *Econometrica Journal of The Econometric Society*, 50(1), 43–61.
- Lee, S., Kang, D., & Kim, M. (2009). Determinants of Crime Incidence in Korea : A Mixed GWR Approach. *World Conference of the Spatial Econometrics Association*, (July), 8–10. Barcelona, Spain.
- Lima, M. L. C., Ximenes, R. A. A., Ramos de Souza, E., Luna, C. F., & Albuquerque, M. de F. P. M. (2005). Análise espacial dos determinantes socioeconômicos dos homicídios no Estado de Pernambuco. *Revista de Saúde Pública*, 39(2), 176–182.
- Lu, B., Harris, P., Charlton, M., & Brunsdon, C. (2014). The GWmodel R package: Further topics for exploring spatial heterogeneity using geographically weighted models. *Geo-Spatial Information Science*, 17(2), 85–101. <https://doi.org/10.1080/10095020.2014.917453>
- Moreira, R. do C., & Fochezatto, A. (2017). Análise Espacial Da Criminalidade No Estado Da Bahia. *RDE - Revista de Desenvolvimento Econômico*, 3(38), 52–80. <https://doi.org/10.21452/rde.v3i38.5033>
- Nakaya, T., Fotheringham, A. S., Brunsdon, C., & Charlton, M. (2005). Geographically weighted Poisson regression for disease association mapping. *Statistics in Medicine*, 24(17), 2695–2717. <https://doi.org/10.1002/sim.2129>
- Nakaya, T., Fotheringham, A. S., Charlton, M., & Brunsdon, C. (2009). Semiparametric geographically weighted generalised linear modelling in GWR 4.0. *10th International Conference on GeoComputation*. Sydney, Australia.
- Nery, M. B., Peres, M. F. T., Cardia, N., Vicentin, D., & Adorno, S. (2012). Regimes espaciais: dinâmica dos homicídios dolosos na cidade de São Paulo entre 2000 e 2008. *Revista Panamericana de Salud Publica/Pan American Journal of Public Health*, 32(6), 405–412.
- O'Brien, R. M. (2007). A caution regarding rules of thumb for variance inflation factors. *Quality and Quantity*, 41(5), 673–690. <https://doi.org/10.1007/s11135->

- Oliveira, C. A. de. (2008). Análise espacial da criminalidade no Rio Grande do Sul. *Revista de Economia*, 34(3), 35–60. <https://doi.org/10.5380/re.v34i3.13824>
- Osgood, D. W. (2000). Poisson-Based Regression Analysis of Aggregate Crime Rates. *Journal of Quantitative Criminology*, 16(1), 21–43. <https://doi.org/10.1023/A:1007521427059>
- Pinto, N. G. M., & Coronel, D. A. (2015). A criminalidade no Brasil: uma análise das evidências empíricas. *Revista Brasileira de Administração Científica*, 6(1), 270–280. <https://doi.org/10.6008/spc2179-684x.2015.001.0017>
- Rybarczyk, G., Maguffee, A., & Kruger, D. (2015). Linking Public Health, Social Capital, and Environmental Stress to Crime Using a Spatially Dependent Model. *Cityscape: A Journal of Policy Development and Research*, 17(1).
- Santos, M. J. do, & Kassouf, A. L. (2008). Estudos Econômicos das Causas da Criminalidade no Brasil : Evidências e Controvérsias. *Revista Economia*, 9(2), 343–372.
- Saraiva, M. V., Conceição, O. C., & França, M. T. A. (2017). Os determinantes da criminalidade nos municípios gaúchos: evidências de um modelo econométrico espacial. *Ensaio FEE*, 38(3), 521–552.
- Sass, K. S., Porsse, A. A., & Silva, E. R. (2016). DETERMINANTES DAS TAXAS DE CRIMES NO PARANÁ: UMA ABORDAGEM ESPACIAL*. *Revista Brasileira de Estudos Urbanos e Regionais*, 10(1), 44–63.
- Stein, R. E., Conley, J. F., & Davis, C. (2016). *The differential impact of physical disorder and collective efficacy: A geographically weighted regression on violent crime*. 81(3), 351–365.
- Vilalta, C. J., & Fondevila, G. (2019). Modeling Crime in an Uptown Neighborhood: The Case of Santa Fe in Mexico City. *Papers in Applied Geography*, 12. <https://doi.org/10.1080/23754931.2018.1554502>
- Vilalta, C. J., & Muggah, R. (2016). What explains criminal violence in mexico city? A test of two theories of crime. *Stability*, 5(1), 1–22. <https://doi.org/10.5334/sta.433>
- Vilalta, C. J., Sanchez, T. W., Fondevila, G., & Ramirez, M. (2019). A descriptive model of the relationship between police CCTV systems and crime. Evidence from

Mexico City. *Police Practice and Research*, 20(2), 105–121.

<https://doi.org/10.1080/15614263.2018.1473770>

Wald, A. (1943). Tests of statistical hypotheses concerning several parameters when the number of observations is large. *Transactions of the American Mathematical Society*, 54, 426–482. <https://doi.org/10.1090/S0002-9947-1943-0012401-3>

Wang, L., Lee, G., & Williams, I. (2019). The spatial and social patterning of property and violent crime in Toronto neighbourhoods: A spatial-quantitative approach. *ISPRS International Journal of Geo-Information*, 8(1). <https://doi.org/10.3390/ijgi8010051>

ANEXO

Figura EE - Gráfico de dispersão entre a taxa de roubo e a população estimada.

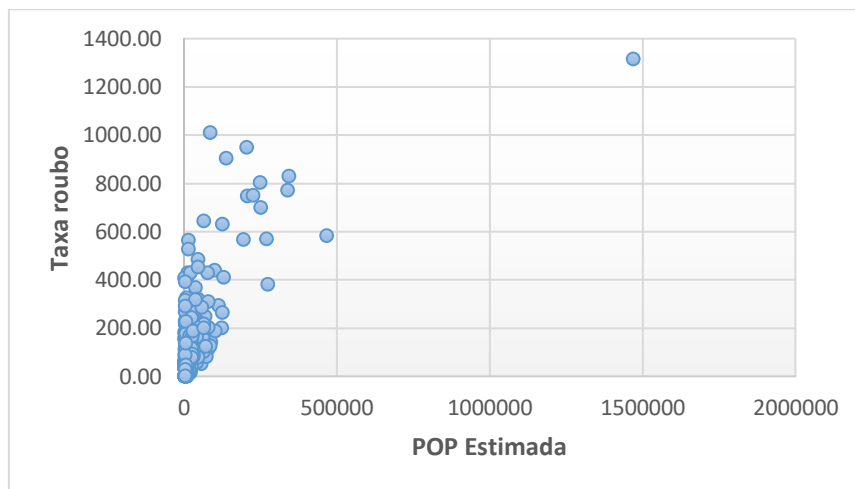


Figura FF - Gráfico de dispersão entre a taxa de roubo e a TAXAHOMEM15A24POP.

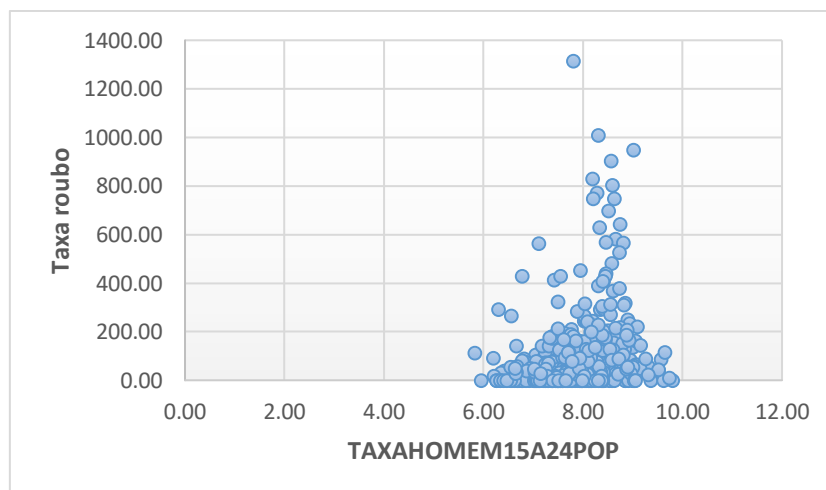


Figura GG - Gráfico de dispersão entre a taxa de roubo e a T_M15A17CF.

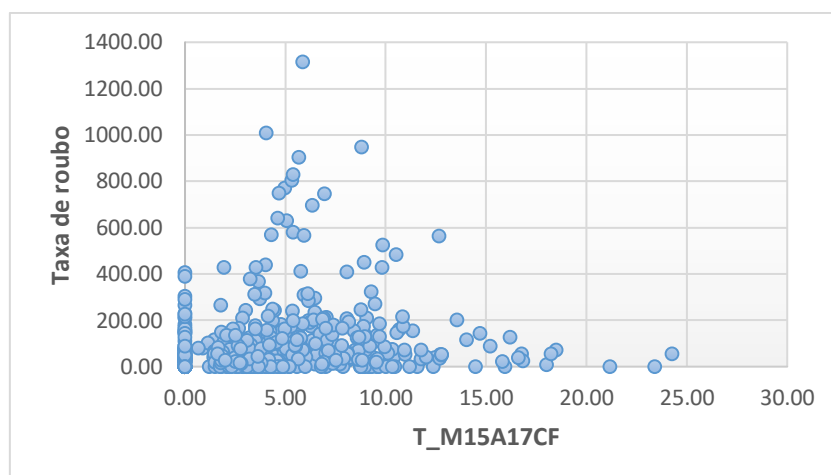


Figura HH - Gráfico de dispersão entre a taxa de roubo e o IDHM.

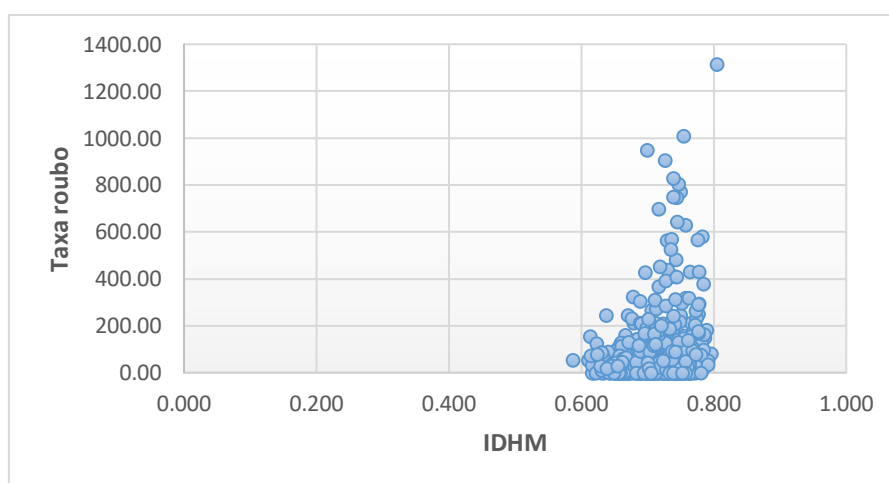


Figura II - Gráfico de dispersão entre a taxa de roubo e a T_OCUPDESLOC_1.

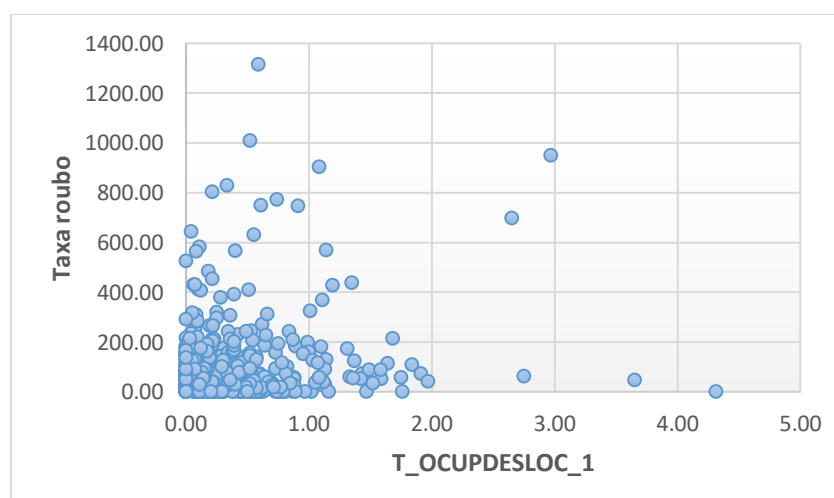


Figura JJ - Gráfico de dispersão entre a taxa de roubo e a T_FUNDIN_TODOS.

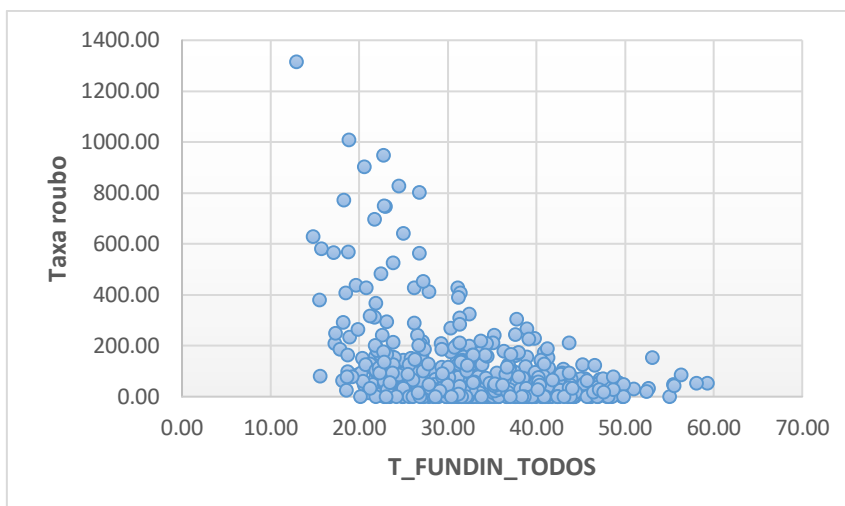


Figura KK - Gráfico de dispersão entre a taxa de roubo e a T_MULCHEFEFIF014.

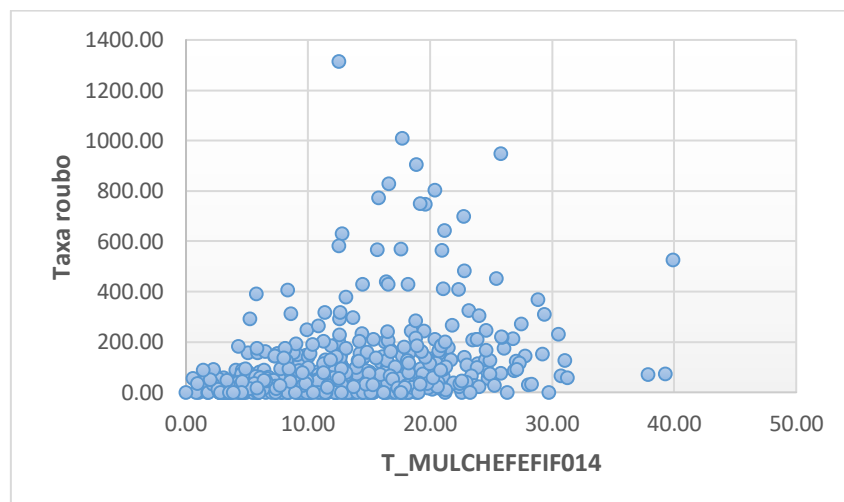


Figura LL - Gráfico de dispersão entre a taxa de roubo e a T_DENS.

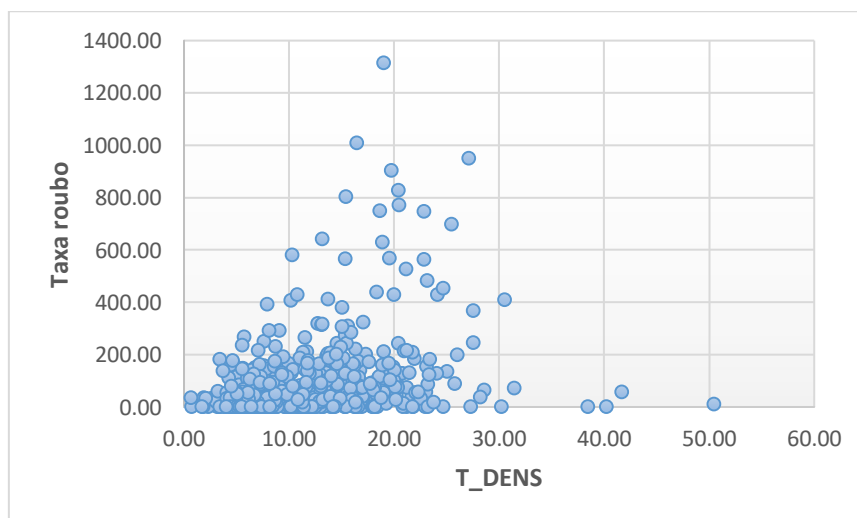


Figura MM - Gráfico de dispersão entre a taxa de roubo e a T_NESTUDA_NTRAB_MMEIO.

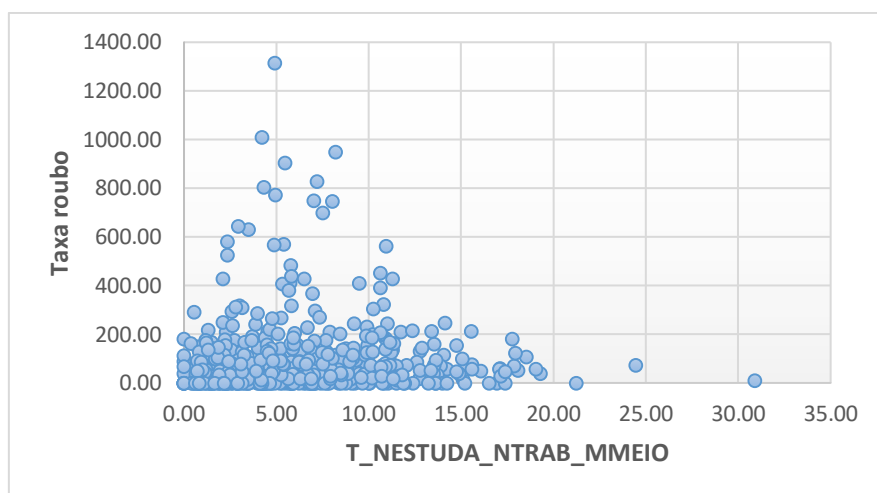


Figura NN - Gráfico de dispersão entre a taxa de roubo e a T_DES18M.

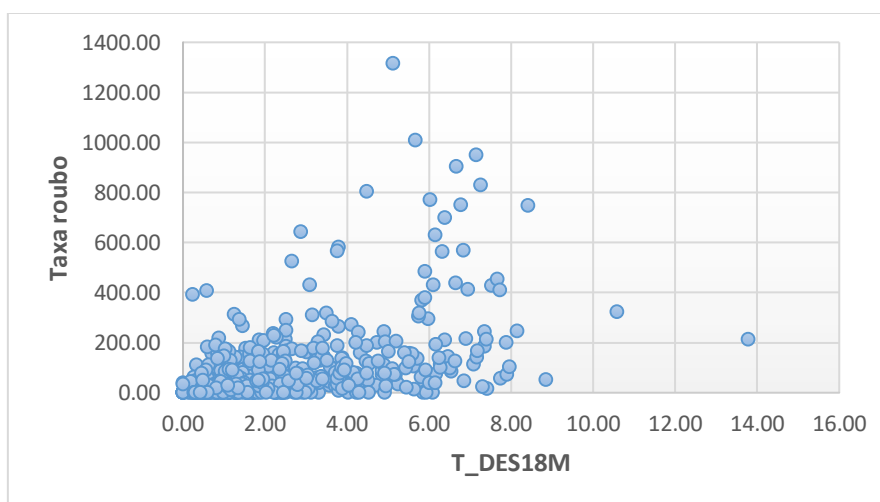


Figura OO - Gráfico de dispersão entre a taxa de roubo e a PIND.

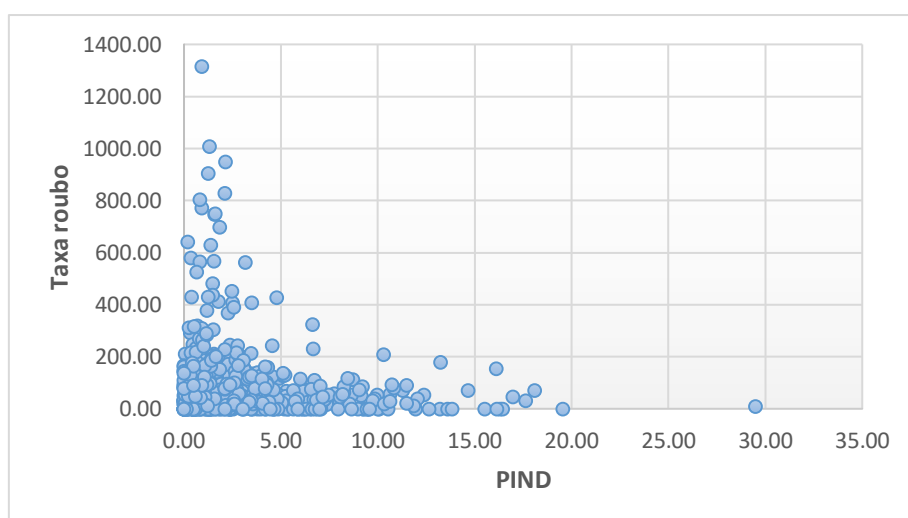


Figura PP - Gráfico de dispersão entre a taxa de roubo o RENOCUP.

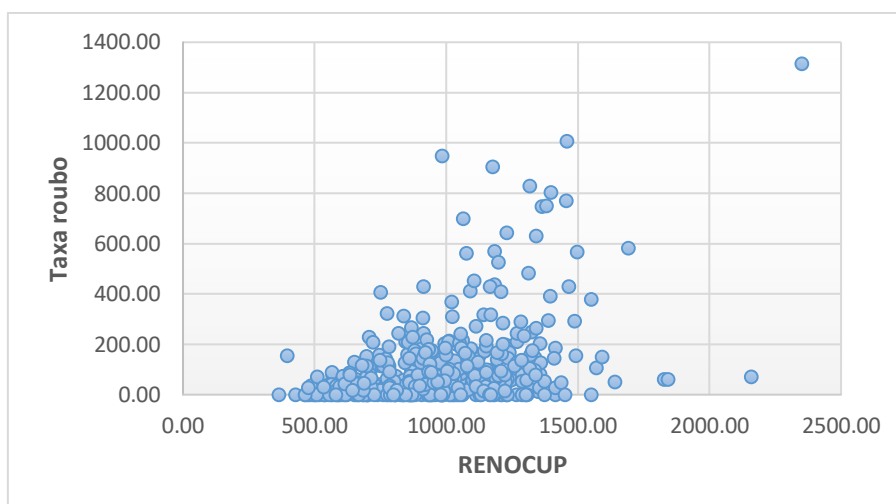


Figura QQ - Gráfico de dispersão entre a taxa de roubo o GINI.

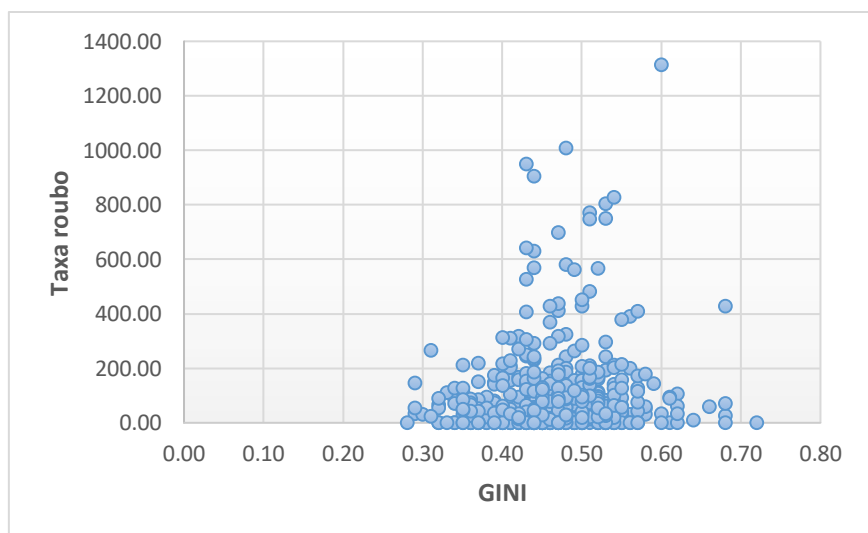


Figura RR - Gráfico de dispersão entre a taxa de roubo e o PREN20RICOS.

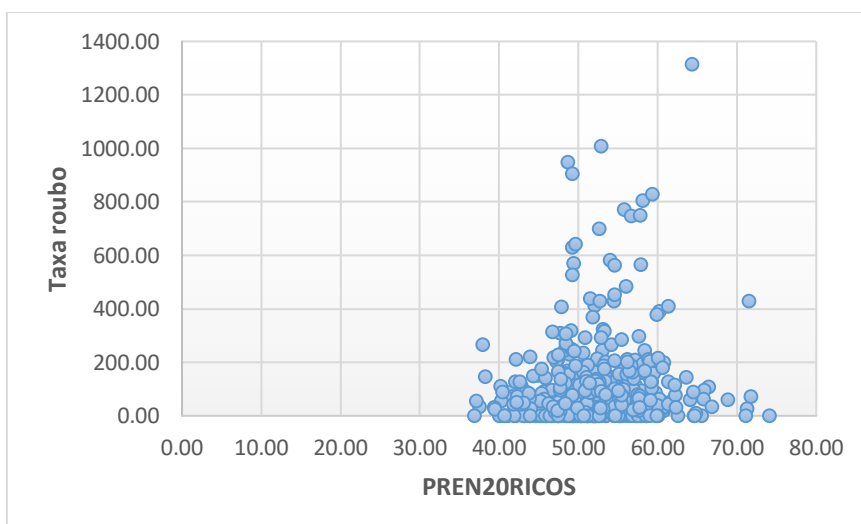


Figura SS - Gráfico de dispersão entre a taxa de roubo e a T_ATRASO_0_MED.

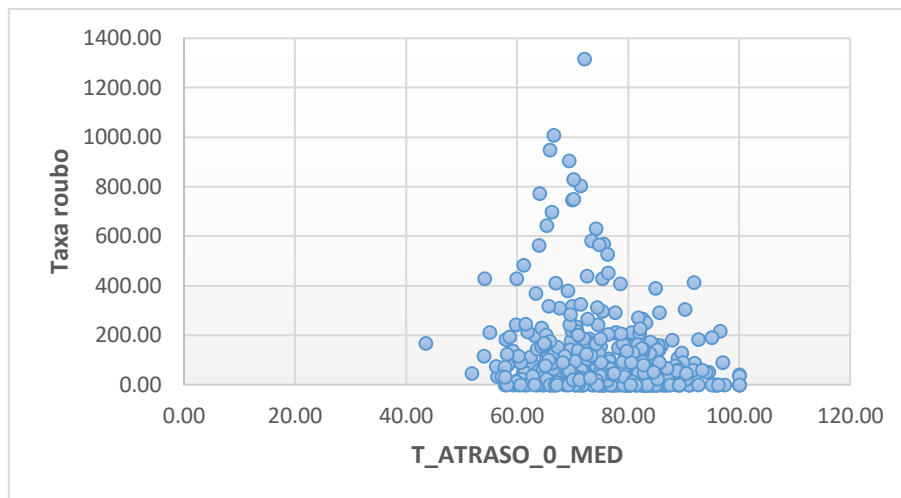


Figura TT - Gráfico de dispersão entre a taxa de roubo e a T_FREQ18A24.

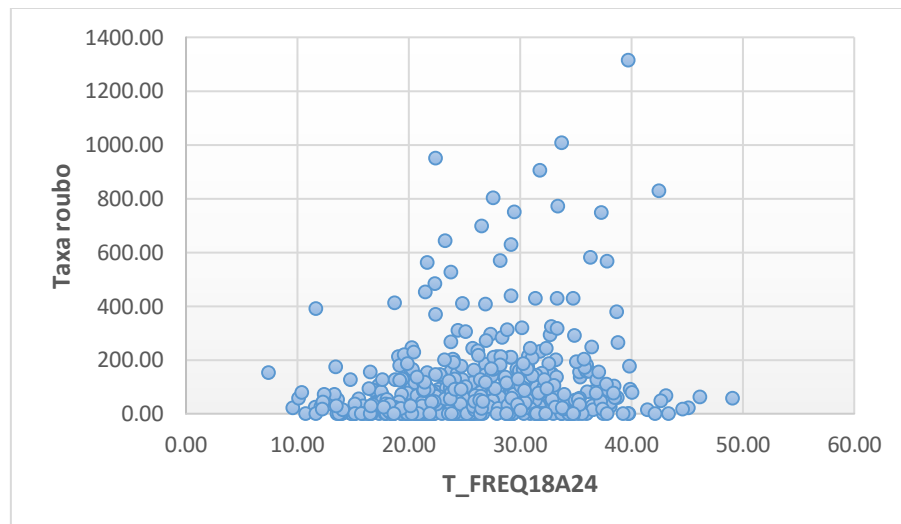


Figura UU - Gráfico de dispersão entre a taxa de roubo e a T_FREQ15A17.

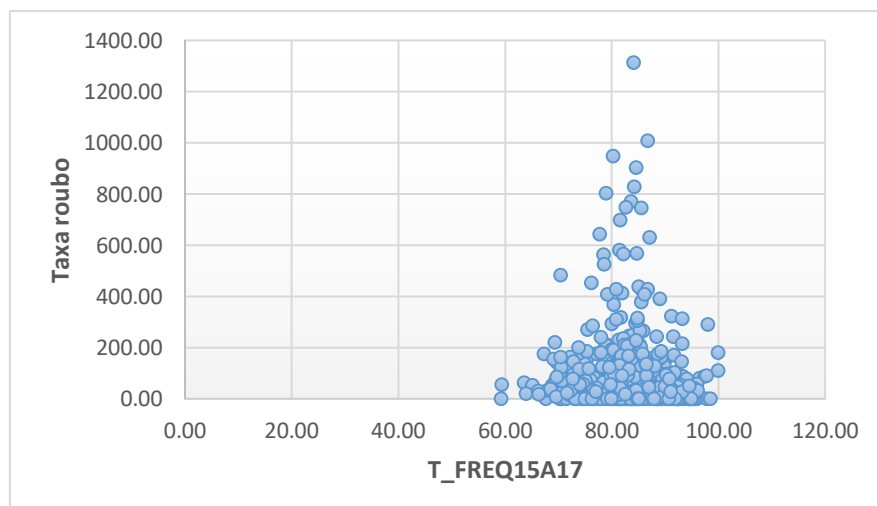


Figura VV - Gráfico de dispersão entre a taxa de roubo e a T_FLMED.

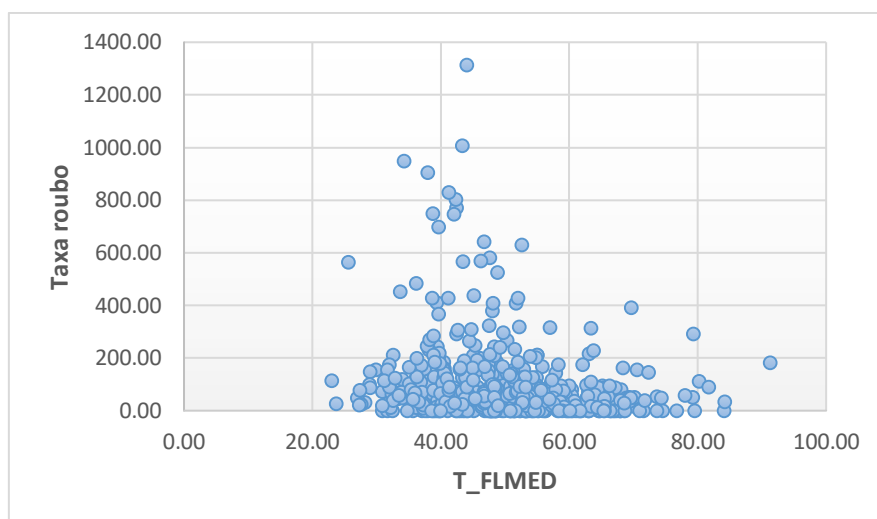


Figura WW - Gráfico de dispersão entre a taxa de roubo e a T_ANALF15M.

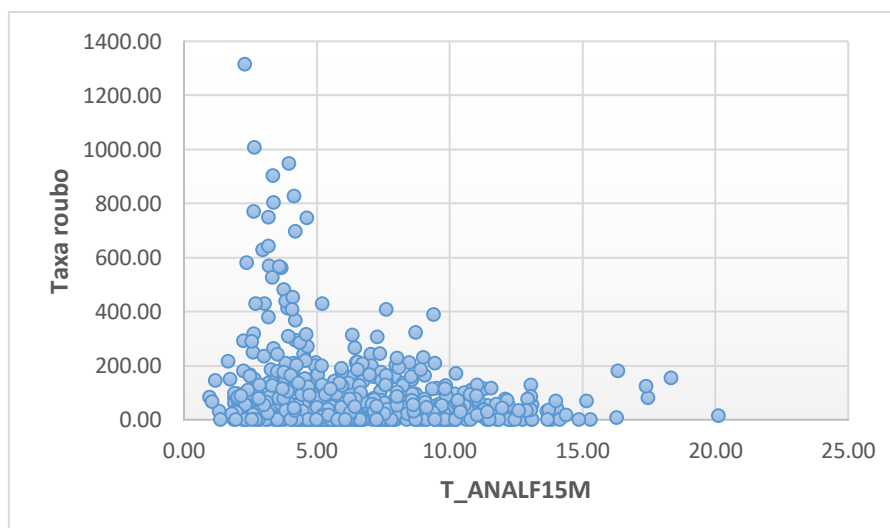


Figura XX - Mapa de *std residual* do modelo GWPR por município no RS.

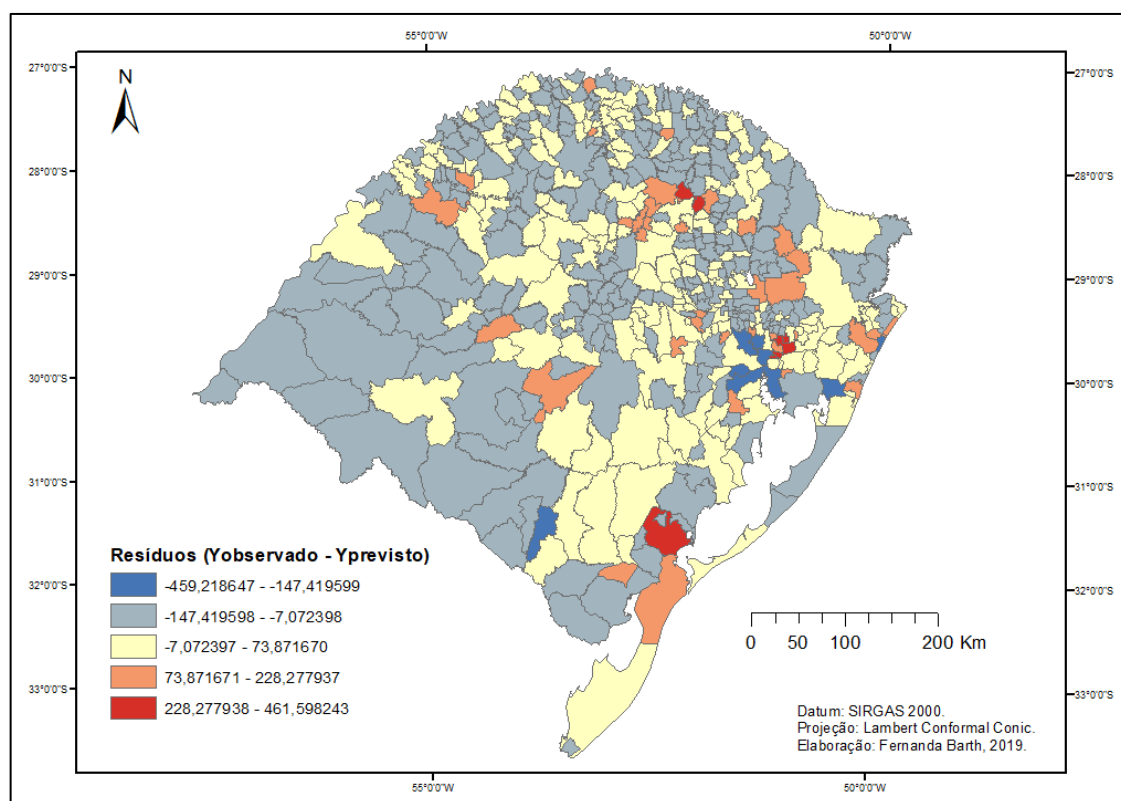


Figura YY - Mapa dos erros padrão para a variável população estimada no modelo GWPR.

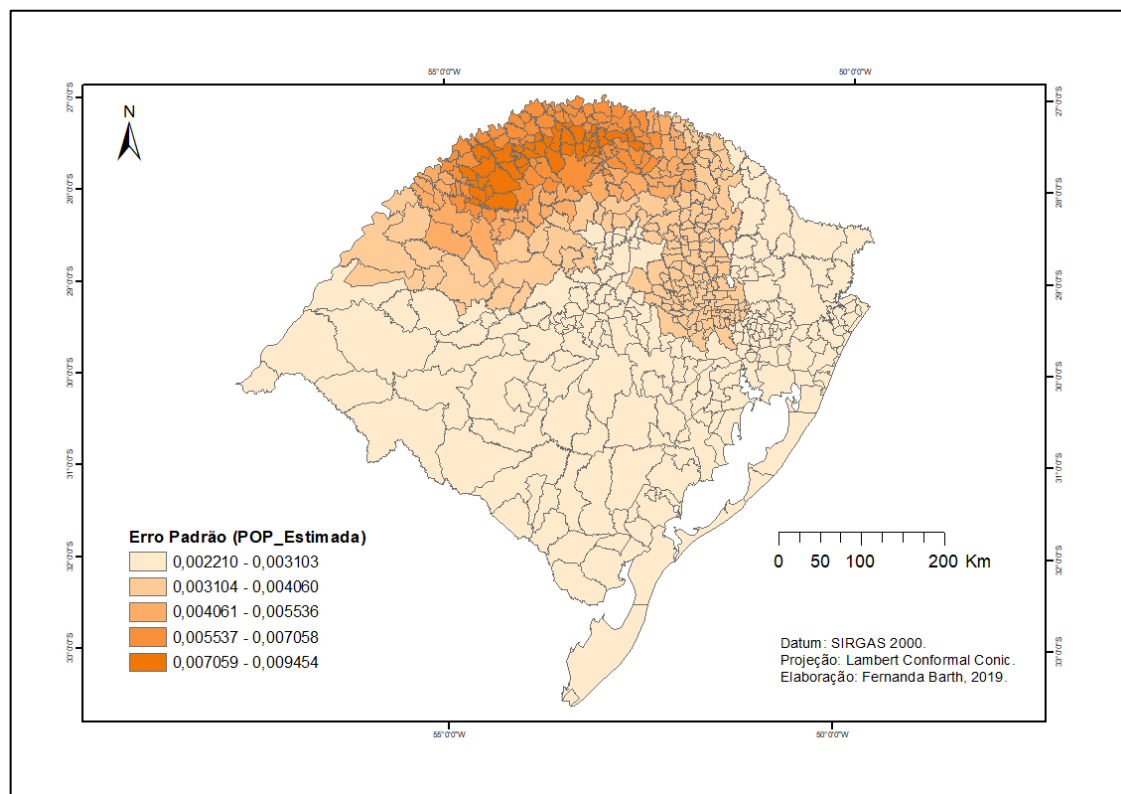


Figura ZZ - Mapa dos erros padrão para a variável PREN20RICO no modelo GWPR.

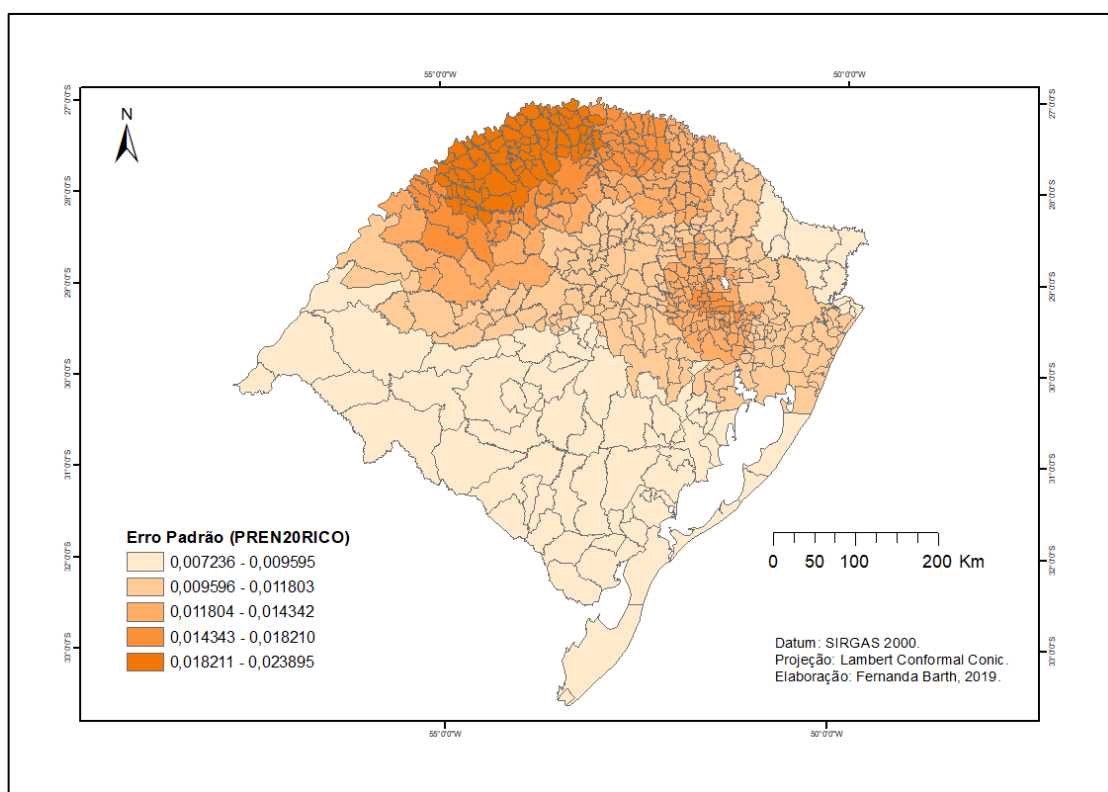


Figura AAA - Mapa dos erros padrão para a variável RENOCUP no modelo GWPR.

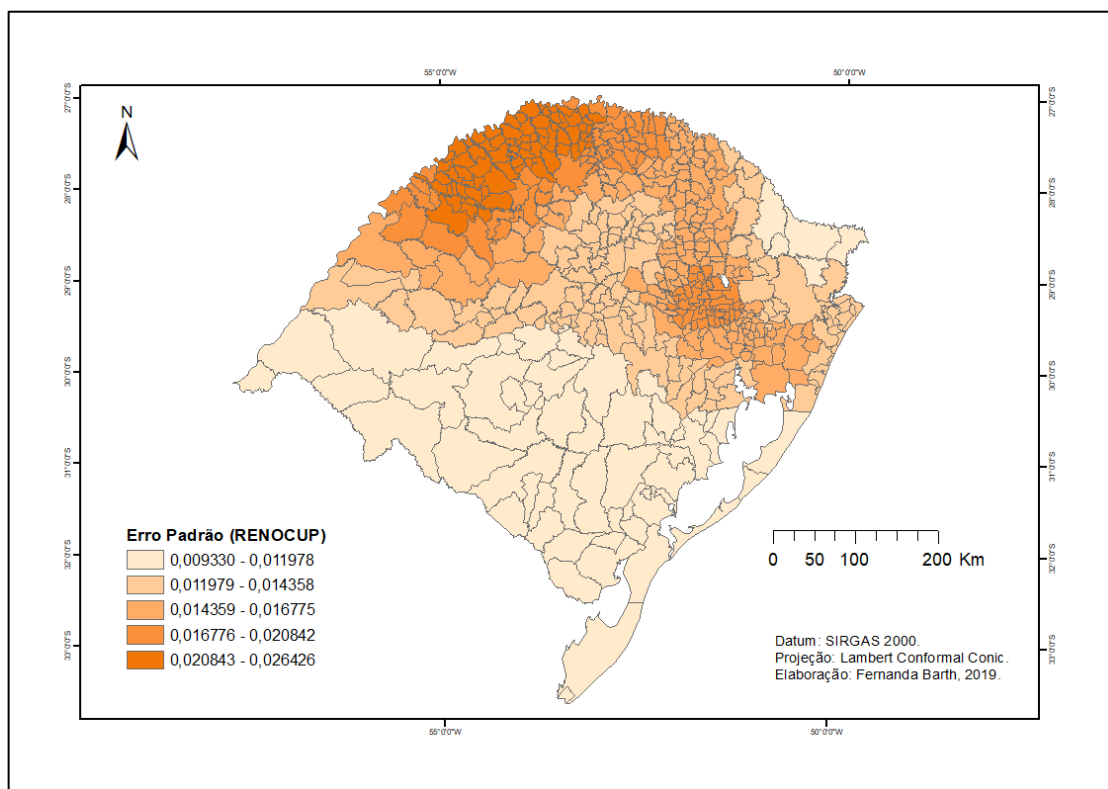


Figura BBB - Mapa dos erros padrão para a variável T_ATRASO_0_MED no modelo GWPR.

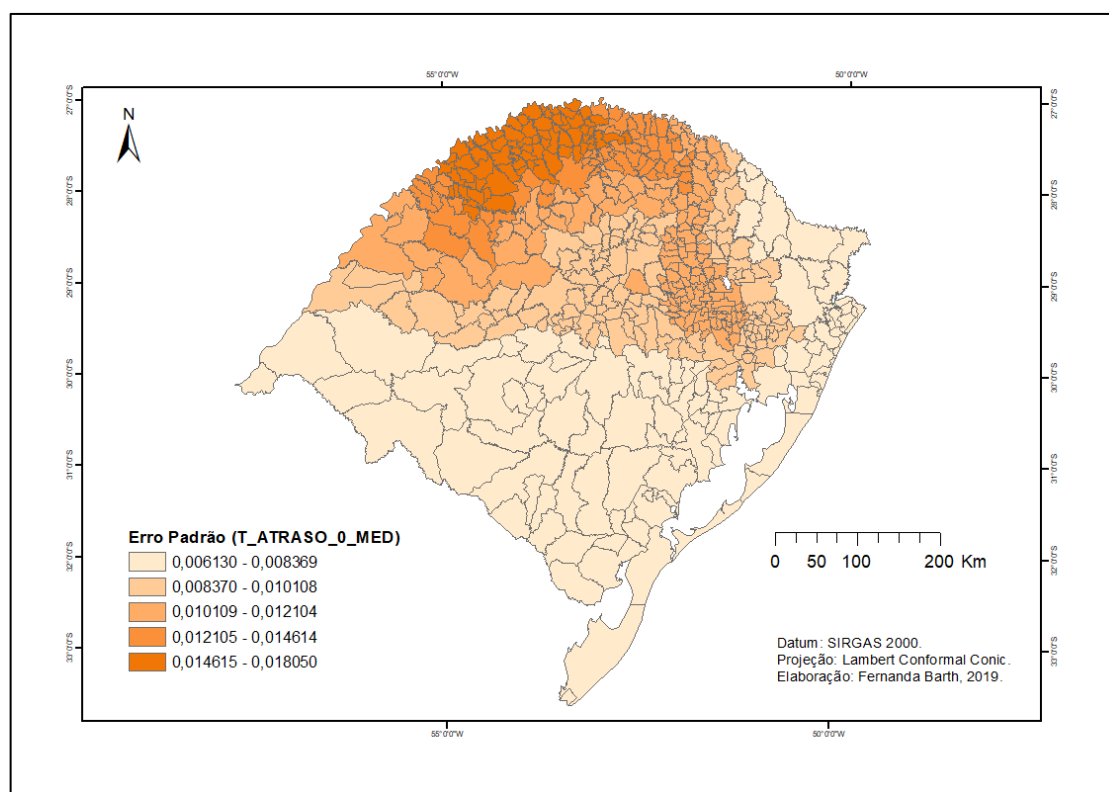


Figura CCC - Mapa dos erros padrão para a variável T_DENS no modelo GWPR.

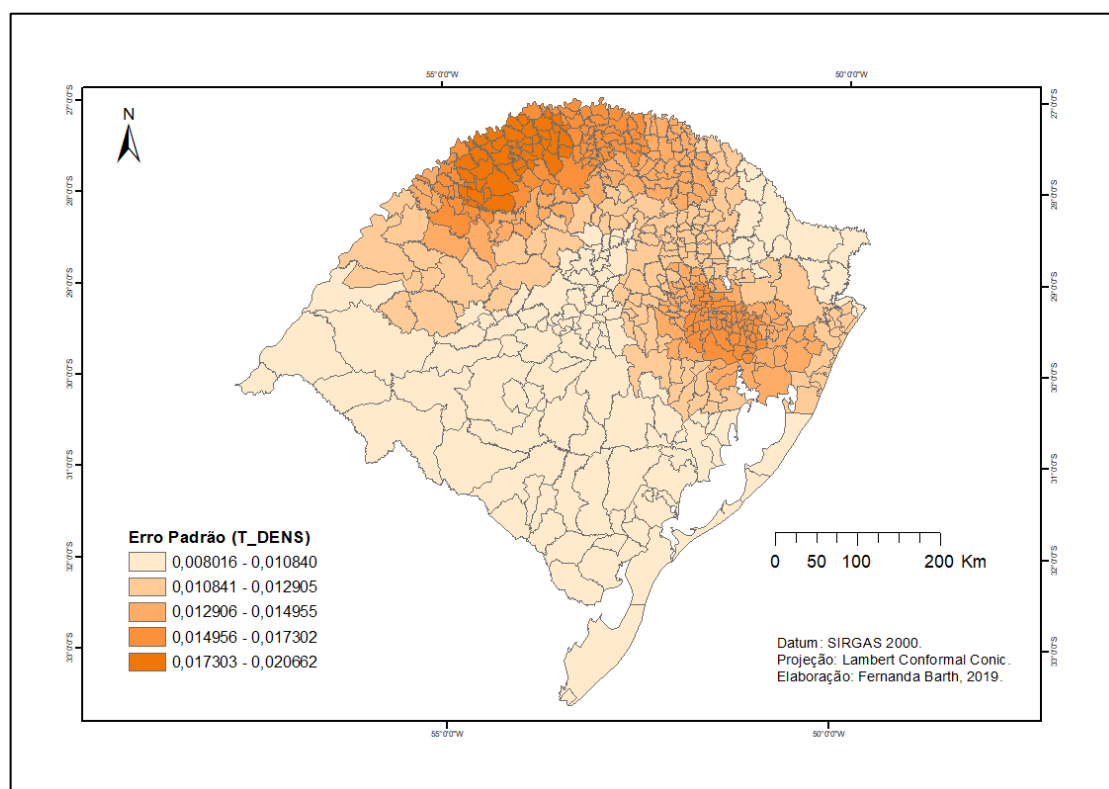


Figura DDD - Mapa dos erros padrão para a variável T_DES18M no modelo GWPR.

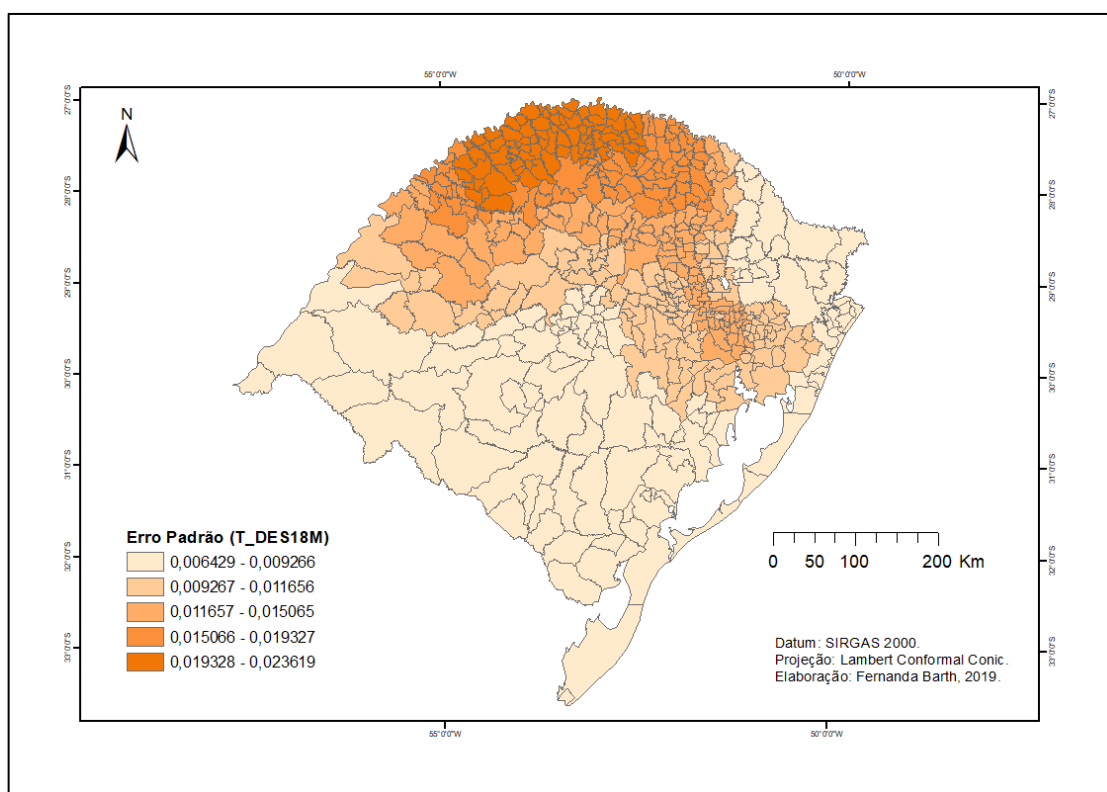


Figura EEE - Mapa dos erros padrão para a variável T_FUNDIN_TODOS no modelo GWPR.

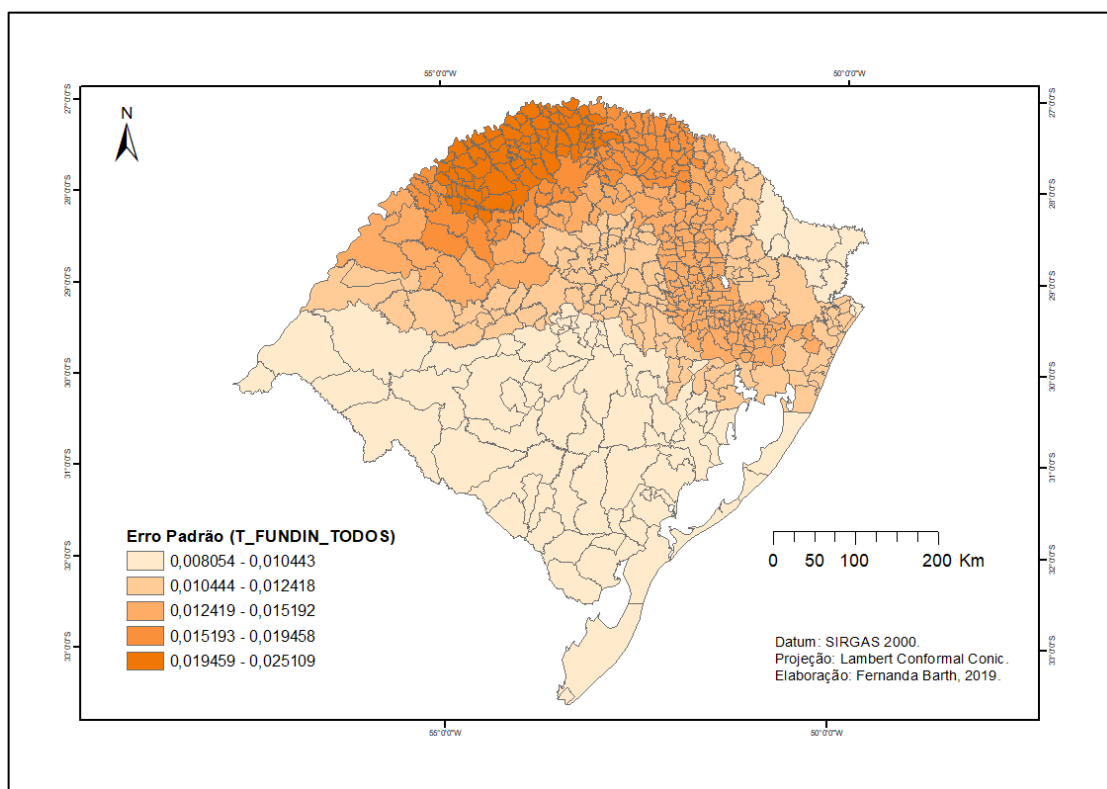


Figura FFF - Mapa dos erros padrão para a variável T_MULCHEFEFIF014 no modelo GWPR.

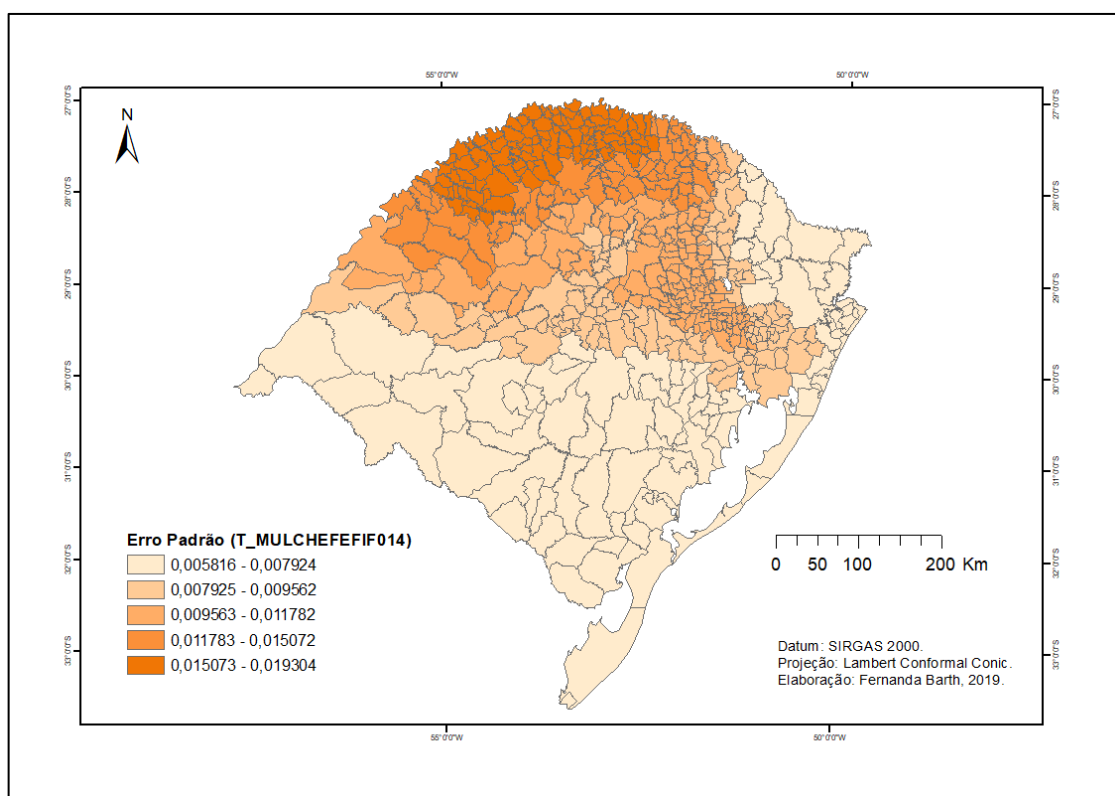
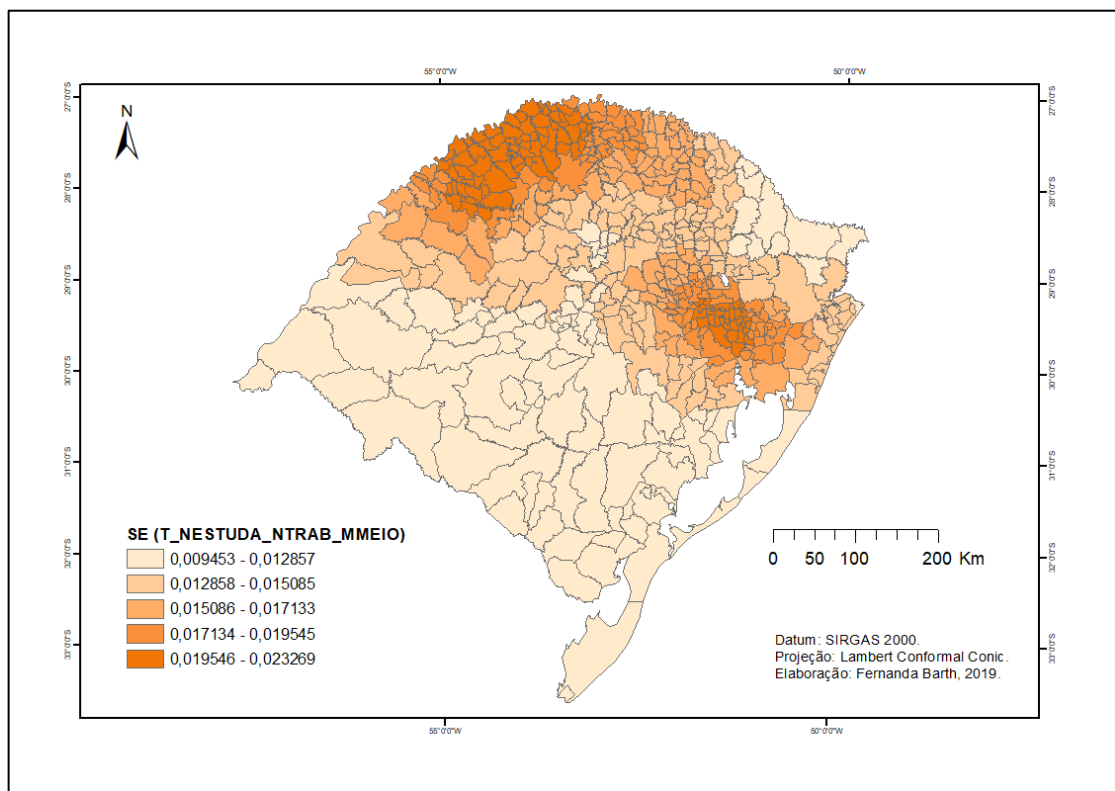


Figura GGG - Mapa dos erros padrão para a variável T_NESTUDA_NTRAB_MMEIO no modelo GWPR.



**C&
SIG**



